Vector Precoding for Gaussian MIMO Broadcast Channels: Impact of Replica Symmetry Breaking

Benjamin M. Zaidel Ralf R. Müller Aris L. Moustakas
Rodrigo de Miguel [1]

February 28, 2011

[1] This work was supported in part by the Research Council of Norway under Grant No. 171133/V30, and by the European Commission under Grant "Newcom++" No. EU-IST-NoE-FP6-2007-216715.

This work was presented in part at the 46th Annual Allerton Conference on Communication, Control & Computing, Monticello, IL, U.S.A, Sep. 2008 and the 10th International Symposium on Spread Spectrum Techniques and Applications (ISSSTA), Bologna, Italy, Aug. 2008.

Benjamin M. Zaidel, Ralf R. Müller and Rodrigo de Miguel are with the Department of Electronics and Telecommunications, The Norwegian University of Science and Technology (NTNU), Trondheim, Norway. e-mail: zaidel@iet.ntnu.no, rodrigo@iet.ntnu.no, ralf@iet.ntnu.no. Aris L. Moustakas is with the Physics Department, National and Kapodistrian University of Athens, Athens, Greece. e-mail: arislm@phys.uoa.gr

Abstract

The so-called "replica method" of statistical physics is employed for the large system analysis of vector precoding for the Gaussian multiple-input multiple-output (MIMO) broadcast channel. The transmitter is assumed to comprise a linear front-end combined with nonlinear precoding, that minimizes the front-end imposed transmit energy penalty. Focusing on discrete complex input alphabets, the energy penalty is minimized by relaxing the input alphabet to a larger alphabet set prior to precoding. For the common discrete lattice-based relaxation, the problem is found to violate the assumption of *replica symmetry* and a *replica symmetry breaking* ansatz is taken. The limiting empirical distribution of the precoder's output, as well as the limiting energy penalty, are derived for one-step replica symmetry breaking. For convex relaxations, replica symmetry is found to hold and corresponding results are obtained for comparison. Particularizing to a "zero-forcing" (ZF) linear front-end, and non-cooperative users, a decoupling result is derived according to which the channel observed by each of the individual receivers can be effectively characterized by the Markov chain u–x–y, where u, x, and y are the channel input, the equivalent precoder output, and the channel output, respectively. For discrete lattice-based alphabet relaxation, the impact of replica symmetry breaking is demonstrated for the energy penalty at the transmitter. An analysis of spectral efficiency is provided to compare discrete lattice-based relaxations against convex relaxations, as well as linear ZF and Tomlinson-Harashima precoding (THP). Focusing on quaternary phase shift-keying (QPSK), significant performance gains of both lattice and convex relaxations are revealed compared to linear ZF precoding, for medium to high signal-to-noise ratios (SNRs). THP is shown to be outperformed as well. In addition, comparing certain lattice-based relaxations for QPSK against a convex counterpart, the latter is found to be superior for low and high SNRs but slightly inferior for medium SNRs in terms of spectral efficiency.

1 Introduction

The multiple-input multiple-output (MIMO) Gaussian broadcast channel (GBC) is the focus of many research activities, addressing the growing demand for higher throughput wireless systems, and in particular the increasing use of multiple-antenna systems in essentially all modern wireless standards (see, e.g., [1–3]). The capacity region of the MIMO GBC is the *dirty paper coding* (DPC) [4] capacity region [5], and several attempts have been made in recent years to propose practically oriented approaches for implementing DPC, as e.g., [6–8]. DPC still remains, however, a difficult, computationally demanding task, which motivates the search for more practical (suboptimum) precoding alternatives.

Since linear precoding, such as zero-forcing (ZF), leads to reduced performance (especially when the channel is ill-conditioned), much attention has been given to nonlinear precoding schemes. In particular, lattice-based precoding approaches have often been investigated, as for example the *vector perturbation* approach suggested in [9] (see also [10] for a general framework). The vector perturbation approach was inspired by the idea of Tomlinson-Harashima precoding (THP) [11] [12]. In this scheme, a scaled *complex integer* vector is added to each data vector, chosen to minimize the energy penalty imposed by a linear zero-forcing (ZF) front-end. A modulo function is employed at the receivers, uniquely determining the transmitted symbols in the absence of noise. An analogous precoding scheme based on a linear minimum-mean-squared-error (MMSE) front-end was considered in [13]. An approach based on optimizing mutual information was taken in [14]. Vector perturbation is however still complex as it involves the solution of an NP-hard integer-lattice least squares problem (commonly implemented using the sphere-decoding algorithm [15]). Addressing the complexity aspect of the method, related approaches can also be found, e.g., in [16, 17] (see references therein for additional literature in this framework), where lattice-basis reduction techniques are employed.

The analytical performance analysis of such nonlinear precoding schemes is not at all trivial. It is common to consider, therefore, uncoded symbol error probabilities (via simulations), asymptotic capacity scaling laws and diversity orders (the asymptotic slope of the error probability in the high signal-to-noise ratio (SNR) regime), or to employ Monte-Carlo simulations to obtain information-theoretically achievable rates (see e.g., [9, 13, 16–19], and also [20] for a semi-tutorial review in this respect). The energy penalty induced by the linear front-end is another commonly addressed performance measure. A lower bound on the energy penalty based on lattice theoretic arguments can be found in [21]. The optimum constellation shaping for a ZF front-end (in terms of the energy penalty), allowing for data to be independently decoded by the users, is investigated in [22], where a selective mapping technique is introduced based on random coding arguments, implementable using nested lattice coding in a trellis precoding framework (see also [23] for a more recent study on selective mapping).

The energy penalty minimization was also investigated in [24] where another nonlinear precoding approach in this framework was recently proposed. The transmitter comprises a linear front-end combined with nonlinear precoding. The nonlinear part relies on relaxation of the transmitted symbols' alphabets to larger alphabet sets. The idea is to optimize the vector of transmitted symbols over the extended alphabet sets, so as to minimize the energy penalty imposed by the linear front-end, which is essentially the idea behind vector perturbation. However, a notable feature of this precoding scheme is that it can also be combined with *convex* extended alphabet sets (in contrast to [9]), lending themselves to *efficient* practical energy minimization algorithms. It can be considered in this sense as a generalization of the vector perturbation scheme (see also [25] in this respect).

Another interesting contribution of [24] is the harnessing of statistical physics tools for the analysis of the nonlinear precoding scheme, while considering the large system limit in which both the number of users and the number of transmit antennas grow large, while their ratio goes to some finite constant. One of the main objectives of statistical physics is the quantitative description of macroscopic properties of many-body systems while starting from the fundamental interactions between microscopic elements. In this framework, a general tool for the analysis of random ("disordered") systems, referred to as the "replica method", was originally invented for the analysis of spin glasses. The latter term describes a spin orientation that has similarity to the type of location of atoms in glasses, which are random in space but frozen in time [26]. However, the replica method turns out to have a much wider range of applications (see, e.g., [26,27] for recent tutorial manuscripts). In recent years, in particular following Tanaka's pioneering work [28], the method has been successfully applied to various problems in wireless communications. The replica method has also been recognized by now as an important tool for information-theoretic analyses in cases where "conventional" random matrix theory does not apply. Although the replica method is heuristic in nature, extensive simulations and exact analytical results in the literature suggest that the replica analysis generally yields excellent approximations in many cases of interest (see again [26,27], and also, e.g., [28–31] and references therein).

The replica analysis usually employs a number of underlying assumptions regarding the behavior of the quantities in concern in the large-system limit. One such fundamental assumption is the "self-averaging" property, which relies on the expectation that macroscopic properties of large random systems converge to deterministic values as the system dimensions grow large. Self-averaging is a property of most physical systems with large (or infinite) degrees of freedom, and is the result of the high probability of occurrence of typical events or samples. Nevertheless, in the case of glassy systems this property has been not trivial to prove, since the underlying randomness of the interactions makes the system inherently non-ergodic. The self-averaging property for the conventional so-called Sherrington-Kirkpatrick (SK) spin glass model [32] was first proven in [33]. More recently, [34, 35] generalized it to a more general class of spin glass models and, using an ingenious method, showed that the averages over the disorder actually do converge in the large system limit. Even more recently, code-division multiple-access (CDMA) systems were shown to be self-averaging [36]. Although the particular system we study is not explicitly covered by the above analysis, it can readily be proved to be self-averaging using the same method. For the sake of space, we will not cover the proof here, however, we will refer to self-averaging as a fundamental property of the large system limit rather than an assumption. Another common assumption in replica analyses is that of *replica symmetry* (RS) (see, e.g., [24, 28, 30]), according to which it is assumed that the crosscorrelations between replicated microscopic system configurations are independent of the replica indices. The RS assumption, however, is known to produce incorrect conclusions for certain physical quantities such as, e.g., the minimum energy configuration. This led to the development of the *replica symmetry breaking* (RSB) theory [26, 37]. Recently, the full RSB solution of the SK spin glass model, first proposed in [38], was shown to be an upper bound to the minimum energy configuration [39] and later to be the exact solution of the model [37]. Apart from its general seminal importance, it is profoundly relevant in the context of vector precoding because the SK-model is a particular case of the more general models discussed in [24] and in the sequel.

In this paper we consider a communication system setting in which RSB indeed occurs, and demonstrate the significant impact of the RSB treatment on the validity of the approximations produced by the replica analysis. We focus here on a wireless MIMO broadcast channel (BC) setting, where the transmitter has N transmit antennas and the K users have single receive antennas. Full channel state information (CSI) is assumed available at the transmitter, while the receivers are cognizant of their own channels only (more on this later). No user-cooperation of any kind is assumed. The received signals are embedded in additive white Gaussian noise (AWGN). The precoding approach considered in [24] is revisited. Note that the focus in [24] is mainly on presenting the method, and on the derivation of the energy penalty in the asymptotic regime, in which both the number of transmit antennas N and the number of users K go to infinity, while $K/N \to \alpha < \infty$ (commonly referred to as the *system load*). Furthermore, the analysis in [24] is based on the RS assumption. It turns out however that the RS assumption can only produce valid asymptotic approximations in this setting when the extended alphabets are convex sets (see, e.g., supportive simulation results in [25]). In contrast, for the non-convex alphabets considered in [24] these approximations turn out to be rather loose, and produce overoptimistic results, especially as the system load gets close to unity. This behavior can be readily observed by comparing the RS based energy penalty to the asymptotic lower bound of [21].

Here, an alternative analysis is provided based on what is referred to in the statistical physics literature as the *one-step RSB (1RSB) ansatz*, which allows one to search for more general solutions than the RS ansatz, but does not cover the full complexity of solutions of full RSB. In addition to an energy penalty analysis, analogous to the one in [24], we complement the results by providing an information-theoretic perspective of the proposed precoding approach. Coded transmissions and achievable throughputs are considered. The employed performance measure is the normalized spectral efficiency, defined as the total number of bits/sec/Hz *per transmit antenna* that can be transmitted arbitrarily reliably through the broadcast channel. The limiting marginal conditional distribution of the nonlinear precoder's output which is required for the calculation of spectral efficiency, as well as the limiting energy penalty, are *analytically formulated.* Focusing on a ZF front-end, the spectral efficiency is expressed via the input-output mutual information of the equivalent single-user channel observed by each of the receivers. The analysis is applied next to a particular family of discrete extended alphabet sets (following [24]), focusing on a QPSK input, demonstrating the RSB phenomenon. To complete the analysis we repeat the derivations while employing the RS ansatz, which is, as said, adequate for convex relaxation schemes, and the results are then applied to a convex alphabet example [24]. For both extended alphabets, numerical spectral efficiency results indicate *significant* performance enhancement over linear ZF preprocessing for medium to high SNRs. Furthermore, performance enhancement is also revealed compared to a generalized THP approach (which is a popular practical nonlinear precoding alternative for such settings). Comparison of the two types of extended alphabet examples leads to interesting conclusions regarding the performance vs. complexity tradeoff of precoding schemes of the kind considered here.

The remainder of this paper is organized as follows. Section 2 describes the system model. Section 3 provides an outline of the replica analysis and includes some general results. In particular, it clarifies the concept of RSB which later results are based upon. In order to analyze the mutual information and later the trade-off between spectral and power efficiency of various precoding schemes, we need to characterize the limiting conditional distribution of the precoder output. This task is solved in Section 4 providing a set of nonlinear equations whose solutions characterize the desired distributions. Section 5 particularizes to the ZF front-end and shows that the channel model can be represented as an equivalent concatenated single-user channel. Then, it derives the spectral efficiency of this equivalent concatenated channel. Section 6 particularizes the results of the previous sections to a discrete lattice-based alphabet relaxation of QPSK. Numerical solutions of the analytical results are provided. Those based on RSB are shown to match simulation results while those based on RS are demonstrated to fail. Section 7 is the corresponding counterpart to Section 6 for convex relaxation. Unlike Section 6, it finds the RS ansatz to provide accurate approximations. Section 8 presents a comparative analysis of the spectral efficiency of the two alphabet relaxation schemes against some other precoding approaches. Finally, Section 9 ends this paper with some concluding remarks.

2 System Model

Consider the following Gaussian MIMO broadcast channel

$$\boldsymbol{r} = \boldsymbol{H}\boldsymbol{t} + \boldsymbol{n} \tag{2-1}$$

where $\boldsymbol{r}_{[K\times 1]}$ is the vector of received signals, $\boldsymbol{H}_{[K\times N]}$ is the (random) complex channel transfer matrix, assumed to be of unit expected row norm, $\boldsymbol{t}_{[N\times 1]}$ is the vector of transmitted signals, and $\boldsymbol{n}_{[K\times 1]}$ is the vector of i.i.d. zero mean proper complex AWGNs at the users' receivers. We denote the noises' spectral levels by σ^2 so that $\boldsymbol{n} \sim \mathcal{N}_c(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$.

The precoding process at the transmitter is depicted in Figure 1. It is assumed that the users' messages are independently encoded, and that the encoders produce coded symbols $\{u_k\}_{k=1}^{K}$ taken from some *discrete* alphabet $\mathscr{U}$. These symbols are treated as random variables, independent across users, and subject to the identical underlying discrete probability $P_U(\tilde{u})$, $\tilde{u} \in \mathscr{U}$. We use

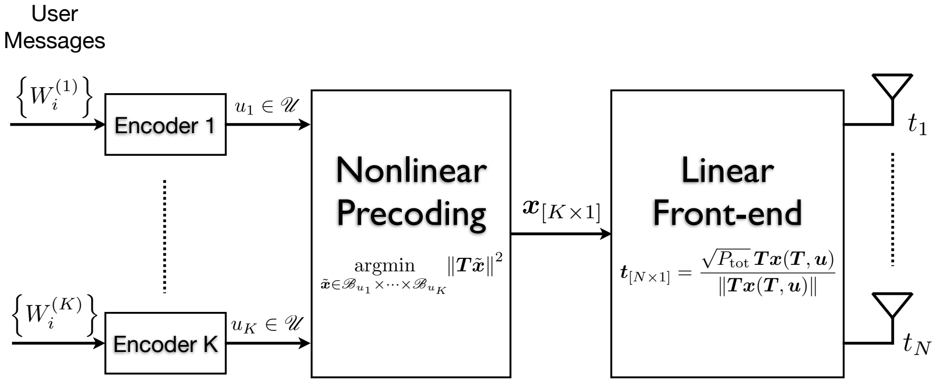
User Messages
$\left\{W_i^{(1)}\right\}$
Encoder 1
$u_1 \in \mathscr{U}$
$\left\{W_i^{(K)}\right\}$
Encoder K
$u_K \in \mathscr{U}$
Nonlinear Precoding
$\underset{\tilde{\boldsymbol{x}} \in \mathscr{B}_{u_1} \times \cdots \times \mathscr{B}_{u_K}}{\operatorname{argmin}} \|\boldsymbol{T}\tilde{\boldsymbol{x}}\|^2$
$\boldsymbol{x}_{[K\times 1]}$
Linear Front-end
$\boldsymbol{t}_{[N\times 1]} = \frac{\sqrt{P_{\text{tot}}}\,\boldsymbol{T}\boldsymbol{x}(\boldsymbol{T},\boldsymbol{u})}{\|\boldsymbol{T}\boldsymbol{x}(\boldsymbol{T},\boldsymbol{u})\|}$
t_1
t_N

Figure 1: Block diagram of the vector precoding scheme.

henceforth for convenience (as shall be made clear in Section 5), the following probability density function (pdf) formulation

$$\mathrm{d}F_U(u) \triangleq f_U(u)\,\mathrm{d}u \triangleq \sum_{\tilde{u}\in\mathscr{U}} P_U(\tilde{u})\delta(u-\tilde{u})\,\mathrm{d}u \quad . \tag{2-2}$$

Let $\boldsymbol{u}_{[K\times 1]}$ denote the vector of the encoders' outputs, i.e., $\boldsymbol{u} = [u_1,\ldots,u_K]^T \in \mathscr{U}^K$. The vector $\boldsymbol{u}$ is the input to a nonlinear precoding block that minimizes the energy penalty of the precoder through input alphabet relaxation (see below), and outputs a $K\times 1$ vector $\boldsymbol{x}$. The vector $\boldsymbol{x}$ is then taken as input to the linear front-end block where it is multiplied by the linear front-end matrix $\boldsymbol{T}_{[N\times K]}$, which is, in general, a function of the channel transfer matrix $\boldsymbol{H}$ (note that $\boldsymbol{x}$ depends on $\boldsymbol{T}$ and, hence, we can use the functional notation $\boldsymbol{x}(\boldsymbol{T},\boldsymbol{u})$). The result is then normalized so that the actually transmitted vector $\boldsymbol{t}$ satisfies an instantaneous *total* power (energy per symbol) constraint P_{tot}, i.e.,

$$\boldsymbol{t} = \sqrt{P_{\mathsf{tot}}}\,\frac{\boldsymbol{T}\boldsymbol{x}(\boldsymbol{T},\boldsymbol{u})}{\|\boldsymbol{T}\boldsymbol{x}(\boldsymbol{T},\boldsymbol{u})\|} \triangleq \sqrt{\frac{P_{\mathsf{tot}}}{\mathscr{E}^{\mathsf{tot}}(\boldsymbol{T},\boldsymbol{x})}}\,\boldsymbol{T}\boldsymbol{x}(\boldsymbol{T},\boldsymbol{u}) \quad , \tag{2-3}$$

where $\mathscr{E}^{\mathsf{tot}}(\boldsymbol{T},\boldsymbol{x})$ denotes the energy penalty induced by the precoding matrix $\boldsymbol{T}$, and the particular choice of $\boldsymbol{x}$, as well as the average symbol energy of the underlying alphabet $\mathscr{U}$ (the explicit dependence on the arguments is omitted henceforth for simplicity). Denoting by P the individual power constraint *per user* (taken as equal for all), so that $P_{\mathsf{tot}} = KP$, we define the *transmit* SNR as

$$\mathsf{snr} \triangleq \frac{P_{\mathsf{tot}}}{K\sigma^2} = \frac{P}{\sigma^2} \quad . \tag{2-4}$$

The energy penalty minimization is performed in the following way. The original alphabet $\mathscr{U}$ is extended ("relaxed") to an alphabet $\mathscr{B} = \bigcup_{\tilde{u}\in\mathscr{U}} \mathscr{B}_{\tilde{u}}$, where the sets $\{\mathscr{B}_{\tilde{u}}\}$ are *disjoint.* The idea here is that every coded symbol $u \in \mathscr{U}$ can be represented *without ambiguity* using any element of $\mathscr{B}_u$ [24][1]. The vector $\boldsymbol{x} = [x_1, \ldots, x_K]^T$ thus satisfies

$$\boldsymbol{x} = \underset{\tilde{\boldsymbol{x}} \in \mathscr{B}_{u_1} \times \cdots \times \mathscr{B}_{u_K}}{\operatorname{argmin}} \|\boldsymbol{T}\tilde{\boldsymbol{x}}\|^2 \quad . \tag{2-5}$$

We note at this point that as an alternative to the normalization taken in (2-3), ensuring an *instantaneous* transmit power constraint, a weaker *average* transmit power constraint can be applied, by simply replacing $\mathscr{E}^{\mathsf{tot}}$ with $E\left\{\mathscr{E}^{\mathsf{tot}}\right\}$, where $E\left\{\cdot\right\}$ denotes expectation. However, since we later concentrate on the energy penalty *per symbol*,

$$\bar{\mathscr{E}} \triangleq \frac{\mathscr{E}^{\mathsf{tot}}}{K} \quad , \tag{2-6}$$

and in view of the self-averaging property of the large system limit (as shall be made clear in the following), the two types of energy constraints yield the same asymptotic results. We thus focus for convenience throughout this paper on the instantaneous power constraint (as implied by (2-3)). Note also that in order to differentiate between the energy penalty induced by the precoding scheme, and the effect of the underlying symbol energy of the input alphabet $\mathscr{U}$, one can alternatively represent the results in terms of what we refer to here as the *precoding efficiency*, defined through

$$\zeta \triangleq \frac{\bar{\mathscr{E}}}{\sigma_u^2} \quad , \tag{2-7}$$

where $\sigma_u^2 = E\{|u|^2\}$ (with the expectation taken with respect to (2-2)).

3 Outline of the Replica Analysis

In the following we describe the main ideas behind the replica analysis of the problem in hand, and provide a heuristic outline of the approach taken to derive the main results of this paper. The reader is referred to tutorial manuscripts such as [26, 27, 31] for an elaborated background on the replica analysis. The fully detailed proofs are deferred to the appendices.

We start here by focusing on the energy penalty, and note that the task of the nonlinear precoding block at the transmitter (see Figure 1) can be described as follows. Its task is equivalent to the minimization of an objective function (called the *Hamiltonian* in physics literature) having the quadratic form

$$\mathcal{H}(\boldsymbol{x}) = \boldsymbol{x}^\dagger \boldsymbol{J} \boldsymbol{x} \quad , \tag{3-1}$$

with $(\cdot)^\dagger$ denoting transpose conjugation and $\boldsymbol{J}$ being a random matrix of dimensions $K \times K$. Thus, the minimum energy penalty per symbol can be expressed as

$$\frac{1}{K} \min_{\boldsymbol{x} \in \mathscr{B}_{\boldsymbol{u}}} \mathcal{H}(\boldsymbol{x}) \quad , \tag{3-2}$$

[1] For practical purposes one would also like to impose additional properties such as a certain minimum distance, although, in principle, the underlying necessary condition is to avoid ambiguity. Note also that the normalization in (2-3) makes the system insensitive to any scaling of the underlying alphabet $\mathscr{U}$.

where we use the shortened notation $\mathscr{B}_{\boldsymbol{u}} \triangleq \mathscr{B}_{u_1} \times \cdots \times \mathscr{B}_{u_K}$. Note also that to comply with (2-5) one should take $\boldsymbol{J} = \boldsymbol{T}^{\dagger}\boldsymbol{T}$, however since the results derived in the sequel hold, at least in part, for a more general class of matrices, we retain the formulation as in (3-1).

To calculate the minimum of the objective function as defined in (3-1), it is convenient to introduce some notions from statistical physics (see, e.g., [31]). In particular, we define a discrete probability distribution on the set of state vectors $\{\boldsymbol{x}\}$, namely the *Boltzmann distribution*, as

$$P_{\mathcal{B}}(\boldsymbol{x}) = \frac{1}{\mathcal{Z}} e^{-\beta \mathcal{H}(\boldsymbol{x})} \tag{3-3}$$

where the parameter $\beta > 0$ is referred to as the *inverse temperature* $\beta = 1/T$, while the normalization factor $\mathcal{Z}$ is the so-called *partition function*, which is defined as

$$\mathcal{Z} = \sum_{\boldsymbol{x} \in \mathscr{B}_{\boldsymbol{u}}} e^{-\beta \mathcal{H}(\boldsymbol{x})} \quad . \tag{3-4}$$

The *energy* of the system is given by

$$\mathcal{E} = \sum_{\boldsymbol{x} \in \mathscr{B}_{\boldsymbol{u}}} P_{\mathcal{B}}(\boldsymbol{x}) \mathcal{H}(\boldsymbol{x}) \quad , \tag{3-5}$$

and the *entropy* (disorder) is defined as

$$\mathcal{S} = - \sum_{\boldsymbol{x} \in \mathscr{B}_{\boldsymbol{u}}} P_{\mathcal{B}}(\boldsymbol{x}) \log P_{\mathcal{B}}(\boldsymbol{x}) \ . \tag{3-6}$$

The definitions above hold for both discrete and continuous alphabets $\mathscr{B}_{\boldsymbol{u}}$. The only difference is that for continuous alphabets the sums over $\boldsymbol{x} \in \mathscr{B}_{\boldsymbol{u}}$ are replaced by integrals.

At thermal equilibrium, the energy of the system is preserved, while the second law of thermodynamics states that the entropy is the maximum possible. This is equivalent to minimizing the *free energy* of the system

$$\mathcal{F} \triangleq \mathcal{E} - \frac{\mathcal{S}}{\beta} \quad , \tag{3-7}$$

where β, the inverse temperature, is in fact the Lagrange multiplier in the maximization of (3-6), subject to the mean energy constraint. At equilibrium, the *free energy* can be expressed as

$$\mathcal{F} = -\frac{1}{\beta} \log \mathcal{Z} \quad . \tag{3-8}$$

Note that from Lagrangian duality the Boltzmann distribution (3-3) is also the solution to the problem of minimizing the energy for a given entropy.

All mean thermodynamic quantities can now be derived directly from the free energy. In particular, the *energy* of the system is

$$\mathcal{E} = \frac{\mathrm{d}(\beta \mathcal{F}(\beta))}{\mathrm{d}\beta} \quad , \tag{3-9}$$

while its thermodynamic *entropy* (disorder) is

$$\mathcal{S} = \beta^2 \frac{\mathrm{d}\mathcal{F}(\beta)}{\mathrm{d}\beta} \quad . \tag{3-10}$$

In addition to the above quantities we can use the free energy and the partition function to obtain the empirical joint distribution of the precoder input u and output x, which is defined for general β as

$$P_{X,U}^{(K)}(\xi, \upsilon) = \frac{1}{K} \sum_{\boldsymbol{x} \in \mathscr{B}_{\boldsymbol{u}}} P_{\mathcal{B}}(\boldsymbol{x}) \sum_{k=1}^{K} 1\left\{(x_k, u_k) = (\xi, \upsilon)\right\} \quad . \tag{3-11}$$

Eqs. (3-3) to (3-11) will be useful in deriving some of the results presented in the sequel.

The rationale behind the introduction of the Boltzmann distribution is that as $\beta \to \infty$, the partition function becomes dominated by the terms corresponding to the *minimum* energy. Hence, taking the logarithm and further normalizing with respect to β, one gets the desired limiting quantity (energy, entropy or empirical distribution) at the minimum energy subspace of $\mathscr{B}_{\boldsymbol{u}}$. Note that even if the energy minimizing vector is not unique, or in fact even if the number of such vectors is exponential in K, one still gets the desired quantity when taking the limit $\beta \to \infty$.

It is crucial to point out that in the above summation over the set of state-vectors $\mathscr{B}_{\boldsymbol{u}}$, both the input vector $\boldsymbol{u}$ and the matrix $\boldsymbol{J}$ are *fixed*. These random variables are called *quenched*. Therefore, all the above manipulations still do not alleviate the difficulty of calculating the desired quantities. In particular, the main difficulty comes from the free energy being a random variable itself, which depends on the particular realizations of $\boldsymbol{J}$ and $\boldsymbol{u}$. As discussed in Section 1, the proofs of the self-averaging property of the SK-model in [33–36] can be generalized to apply to the form of $\mathcal{H}(\boldsymbol{x})$ analyzed here. This means that the free energy converges in probability at the asymptotic limit to a non-random quantity, i.e.,

$$\lim_{K\to\infty} \Pr\left(\frac{1}{K}\left|\mathcal{F} - E\left\{\mathcal{F}\right\}\right| > \epsilon\right) = 0 \quad \forall \epsilon > 0 \quad , \tag{3-12}$$

where the expectation $E\{\cdot\}$ is over all realizations of $\boldsymbol{J}$ and $\boldsymbol{u}$. As a result, all quantities that can be obtained from the free energy in an analytic manner, e.g., by differentiation of a parameter, are also self-averaging. The empirical joint distribution of the precoder input and output converges to a non-random distribution which is expressed by (3-16). This self-averaging property makes the problem more straightforward to tackle, since we may now hope to get analytic results for the average of the free energy and its derivatives.

With that in mind, the limiting energy penalty (per symbol) can be represented as

$$\begin{aligned} \bar{\mathscr{E}} &= \lim_{K\to\infty} \frac{1}{K} \min_{\boldsymbol{x}\in\mathscr{B}_{\boldsymbol{u}}} \boldsymbol{x}^{\dagger}\boldsymbol{J}\boldsymbol{x} = -\lim_{K\to\infty}\lim_{\beta\to\infty} \frac{1}{\beta K} E\left\{\log \sum_{\boldsymbol{x}\in\mathscr{B}_{\boldsymbol{u}}} \mathrm{e}^{-\beta \boldsymbol{x}^{\dagger}\boldsymbol{J}\boldsymbol{x}}\right\} \\ &= \lim_{K\to\infty}\lim_{\beta\to\infty} E\left\{\frac{\mathcal{F}(\beta)}{K}\right\} \quad . \end{aligned} \tag{3-13}$$

To obtain the empirical joint distribution $P_{X,U}$, we follow a technique very common in the physics literature [27], and introduce a dummy variable $h \in \mathbb{R}$, as well as the function

$$V(h, \xi, \upsilon, \boldsymbol{x}, \boldsymbol{u}) = -h \sum_{k=1}^{K} 1\left\{(x_k, u_k) = (\xi, \upsilon)\right\} \quad . \tag{3-14}$$

If we add this term to $\mathcal{H}(\boldsymbol{x})$ in the exponent, the partition function gets modified to

$$\mathcal{Z}(h) = \sum_{\boldsymbol{x} \in \mathscr{B}_{\boldsymbol{u}}} \mathrm{e}^{-\beta(\mathcal{H}(\boldsymbol{x}) + V(h))} \tag{3-15}$$

where we have dropped the explicit dependence of $\mathcal{Z}(h)$ and $V(h)$ on υ and ξ (as well as $\boldsymbol{u}$ and $\boldsymbol{x}$) for the sake of notational compactness. In the sequel, any dependence on h shall implicitly also indicate a dependence on υ and ξ. Using the above partition function we obtain a modified free energy using (3-8). Upon differentiation with respect to h, setting $h = 0$, and letting $\beta \to \infty$ we get

$$P_{X,U}(\xi, \upsilon) = \lim_{K \to \infty} P_{X,U}^{(K)}(\xi, \upsilon) \tag{3-16}$$

$$= \lim_{K \to \infty} \frac{1}{K} \lim_{\beta \to \infty} E\left\{ \left. \frac{\partial \mathcal{F}(\beta, h)}{\partial h} \right|_{h=0} \right\} \tag{3-17}$$

where $\mathcal{F}(\beta, h)$ denotes the free energy for the modified partition function $\mathcal{Z}(h)$.[2]

The next step in the analysis is to invoke some underlying assumptions. The first assumption is that the random matrix $\boldsymbol{J}$ can be decomposed as

$$\boldsymbol{J} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\dagger} \quad , \tag{3-18}$$

where $\boldsymbol{D}$ is a diagonal matrix with diagonal elements being the eigenvalues of $\boldsymbol{J}$, and $\boldsymbol{U}$ is a unitary Haar distributed matrix [40]. It is further assumed that the empirical distribution of the diagonal elements of $\boldsymbol{D}$ converges to a nonrandom distribution uniquely characterized by its R-transform[3] $R(\cdot)$, which is assumed to exist.

Going back to the original communication system model, note that we are in fact interested in the normalized averages of most of the quantities described above, at the limit as $K \to \infty$. Therefore, to make a distinction, while retaining the relation between the quantities, we shall use

[2] An alternative method for deriving the limiting empirical distribution, which relies on the limiting moments, can be found in [30], albeit with more restrictive assumptions on the limiting distribution.

[3] For a definition of the R-transform, see Appendix E.

henceforth the following notational convention

$$\begin{aligned}\mathscr{F}(\beta) &\triangleq \lim_{K\to\infty}\frac{1}{K}E\left\{\mathcal{F}\right\} \quad , && \text{(3-19)}\\ \mathscr{S}(\beta) &\triangleq \lim_{K\to\infty}\frac{1}{K}E\left\{\mathcal{S}\right\} \quad , && \text{(3-20)}\\ \mathscr{E}(\beta) &\triangleq \lim_{K\to\infty}\frac{1}{K}E\left\{\mathcal{E}\right\} \quad . && \text{(3-21)}\end{aligned}$$

Calculating the expectation of a logarithm of a sum of exponents (see (3-13)) is a formidable task. The standard approach in statistical physics is to invoke the so-called replica “trick”. The latter is based on the following identity[4]

$$E\left\{\log\mathcal{Z}\right\} = \lim_{n\to 0^+}\frac{\log E\left\{\mathcal{Z}^n\right\}}{n} \tag{3-22}$$

which holds in general for *real* n. The “trick” here relies on the *assumption* that the right hand side (RHS) of (3-22) can be evaluated for *integer* n, and that the desired quantity can be found by analytic continuation in the vicinity of $n = 0^+$. Although this “trick” does not *a priori* have any justified validity, its success in statistical physics, and more recently in communications theory, makes it a reasonable approach. Further assuming that the limits with respect to K and n can be interchanged (which is the common practice in replica analyses), (3-13) can be rewritten as

$$\begin{aligned}\bar{\mathscr{E}} &= \lim_{\beta\to\infty}\mathscr{E}(\beta)\\ &= -\lim_{\beta\to\infty}\lim_{n\to 0^+}\frac{1}{n}\lim_{K\to\infty}\frac{1}{\beta K}\log E\left\{\left(\sum_{\boldsymbol{x}\in\mathscr{B}_{\boldsymbol{u}}}\mathrm{e}^{-\beta\boldsymbol{x}^\dagger\boldsymbol{J}\boldsymbol{x}}\right)^n\right\}\\ &= -\lim_{\beta\to\infty}\frac{1}{\beta}\lim_{n\to 0}\frac{1}{n}\lim_{K\to\infty}\frac{1}{K}\log E\left\{\sum_{\{\boldsymbol{x}_a\}}\mathrm{e}^{\sum_{a=1}^n -\beta\boldsymbol{x}_a^\dagger\boldsymbol{J}\boldsymbol{x}_a}\right\}\\ &= -\lim_{\beta\to\infty}\frac{1}{\beta}\lim_{n\to 0}\frac{1}{n}\lim_{K\to\infty}\frac{1}{K}\log E\left\{\sum_{\{\boldsymbol{x}_a\}}\mathrm{e}^{-\operatorname{Tr}(\beta\boldsymbol{J}\sum_{a=1}^n\boldsymbol{x}_a\boldsymbol{x}_a^\dagger)}\right\} \quad ,\end{aligned} \tag{3-23}$$

where we use the notation $\sum_{\{\boldsymbol{x}_a\}} = \sum_{\boldsymbol{x}_1\in\mathscr{B}_{\boldsymbol{u}}}\cdots\sum_{\boldsymbol{x}_n\in\mathscr{B}_{\boldsymbol{u}}}$, and $\operatorname{Tr}(\cdot)$ denotes the trace operator.

The summation over the replicated precoder output vectors $\{\boldsymbol{x}_a\}_{a=1}^n$ in (3-23) is performed by splitting the replicas into subshells, defined through an $n\times n$ matrix $\boldsymbol{Q}$

$$S(\boldsymbol{Q}) \triangleq \left\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n \middle| \boldsymbol{x}_a^\dagger\boldsymbol{x}_b = KQ_{ab}\right\}. \tag{3-24}$$

The limit $K\to\infty$ allows us to perform the following derivations by saddle point integration. This first yields the following general result.

Proposition 3.1 *For any inverse temperature* β*, any structure of* $\boldsymbol{Q}$ *consistent with* (3-24)*, and*

[4] An equivalent representation often encountered in the literature is $E\left\{\log\mathcal{Z}\right\} = \lim_{n\to 0^+}\frac{E\{\mathcal{Z}^n\}-1}{n}$.

any R-transform $R(\cdot)$ such that $R(\boldsymbol{Q})$ is well-defined[5]*, the energy is given by*

$$\mathscr{E}(\beta)=\lim_{n\to 0}\frac{1}{n}\operatorname{Tr}\left[\boldsymbol{Q}\,R(-\beta\boldsymbol{Q})\right] \quad , \tag{3-25}$$

where $\boldsymbol{Q}$ is the solution to the saddle point equation

$$\boldsymbol{Q}=\int\frac{\sum\limits_{\mathbf{x}\in\mathscr{B}_u^n}\mathbf{x}\mathbf{x}^\dagger \mathrm{e}^{-\beta\mathbf{x}^\dagger R(-\beta\boldsymbol{Q})\mathbf{x}}}{\sum\limits_{\mathbf{x}\in\mathscr{B}_u^n}\mathrm{e}^{-\beta\mathbf{x}^\dagger R(-\beta\boldsymbol{Q})\mathbf{x}}}\,\mathrm{d}F_U(u) \tag{3-26}$$

with $\mathscr{B}_u^n$ denoting the n-fold Cartesian product of $\mathscr{B}_u$.

Proof: See Appendix C. ■

With the help of Proposition 3.1, the energy can be written as

$$\mathscr{E}(\beta)=\lim_{K\to\infty}\frac{1}{K}\frac{\sum\limits_{\boldsymbol{x}\in\mathscr{B}_{\boldsymbol{u}}}\boldsymbol{x}^\dagger\boldsymbol{J}\boldsymbol{x}\,\mathrm{e}^{-\beta\boldsymbol{x}^\dagger\boldsymbol{J}\boldsymbol{x}}}{\sum\limits_{\boldsymbol{x}\in\mathscr{B}_{\boldsymbol{u}}}\mathrm{e}^{-\beta\boldsymbol{x}^\dagger\boldsymbol{J}\boldsymbol{x}}} \tag{3-27}$$

$$=\lim_{n\to 0}\frac{1}{n}\int\frac{\sum\limits_{\mathbf{x}\in\mathscr{B}_u^n}\mathbf{x}^\dagger R(-\beta\boldsymbol{Q})\mathbf{x}\,\mathrm{e}^{-\beta\mathbf{x}^\dagger R(-\beta\boldsymbol{Q})\mathbf{x}}}{\sum\limits_{\mathbf{x}\in\mathscr{B}_u^n}\mathrm{e}^{-\beta\mathbf{x}^\dagger R(-\beta\boldsymbol{Q})\mathbf{x}}}\,\mathrm{d}F_U(u) \tag{3-28}$$

with $\boldsymbol{Q}$ given by (3-26). In (3-27), $\boldsymbol{x}$ is a K-dimensional vector and its components represent users. The contributions of the users to the energy arise due to the inner product $\boldsymbol{x}^\dagger\boldsymbol{J}\boldsymbol{x}$ and are coupled, unless $\boldsymbol{J}$ is diagonal. In (3-28), $\mathbf{x}$ is an n-dimensional vector and its components represent replicas of the *same* user. The contributions of the users to the energy arise due to integration over the distribution $F_U(u)$, and are decoupled and additive. This is just another incarnation of the decoupling principle that, under the assumption of replica symmetry, was addressed in [30]. Here, we find that it holds for the energy of general (also replica symmetry breaking) spin glass systems and their equivalents in communication theory.

Another interesting observation is the following. In [41], an analogy between the R-transform and effective interference in linear MMSE detection was discovered, and the additivity of the effective interference of coupled users was explained based on the additivity of the R-transforms of free random variables. Relying on the code symbols of different users $\{u_k\}$ being i.i.d., we can rewrite (3-28) as

$$\mathscr{E}(\beta)=\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\underbrace{\lim_{n\to 0}\frac{1}{n}\frac{\sum\limits_{\mathbf{x}\in\mathscr{B}_{u_k}^n}\mathbf{x}^\dagger R(-\beta\boldsymbol{Q})\mathbf{x}\,\mathrm{e}^{-\beta\mathbf{x}^\dagger R(-\beta\boldsymbol{Q})\mathbf{x}}}{\sum\limits_{\mathbf{x}\in\mathscr{B}_{u_k}^n}\mathrm{e}^{-\beta\mathbf{x}^\dagger R(-\beta\boldsymbol{Q})\mathbf{x}}}}_{\mathscr{E}_k(\beta)} \tag{3-29}$$

[5] Note that if $R(\cdot)$ has a series expansion, $R(\boldsymbol{Q})$ is well-defined. Since $R(\cdot)$ is the free cumulant generating function, $R(\boldsymbol{Q})$ is well-defined, if all moments of the asymptotic eigenvalue distribution of $\boldsymbol{J}$ exist.

and interpret $\mathscr{E}_k(\beta)$ as the effective energy of user k. Like the effective interference in [41], it depends only on the signal constellation of user k and the R-transform, and it is additive among users. In contrast to [41], (3-29) is more general and neither constrained to linear detectors nor to Gaussian symbol alphabets.

To produce explicit results, note that the limit $n \to 0$ of the $n \times n$ matrix $\boldsymbol{Q}$ can only be defined imposing a certain structure onto the crosscorrelation matrix $\boldsymbol{Q}$ at the saddle point, unless the summations in (3-26) can be evaluated explicitly, e.g., for $\mathscr{B}_u = \mathbb{C}$. The simplest structure is that of replica symmetry, which in the current setting boils down to

$$\boldsymbol{Q} = q_0 \mathbf{1}_{n\times n} + \frac{\chi_0}{\beta} \boldsymbol{I}_{n\times n} \quad , \tag{3-30}$$

for some constants $\{q_0, \chi_0\}$. The 1RSB assumption leads to a more involved structure, formulated as

$$\boldsymbol{Q} = q_1 \mathbf{1}_{n\times n} + p_1 \boldsymbol{I}_{\frac{n\beta}{\mu_1} \times \frac{n\beta}{\mu_1}} \otimes \mathbf{1}_{\frac{\mu_1}{\beta} \times \frac{\mu_1}{\beta}} + \frac{\chi_1}{\beta} \boldsymbol{I}_{n\times n} \quad , \tag{3-31}$$

using the constants $\{q_1, p_1, \chi_1, \mu_1\}$. The above constants (i.e., $\{q_0, \chi_0\}$ for RS, and $\{q_1, p_1, \chi_1, \mu_1\}$ for 1RSB) are referred to as *macroscopic* parameters, and obtained from the corresponding saddle point equations. The limiting energy penalty can then be expressed in terms of these macroscopic parameters, as shown in the following sections. An analogous procedure can be employed to obtain the limiting empirical joint distribution of the precoder input and output using (3-16).

Replica symmetry breaking is not limited to one step, and in fact in order to *exactly* characterize the limiting energy penalty and precoder output statistics, we would eventually need to consider full RSB, as discussed in Section 1. However, we will only present here precoding results up to the accuracy of 1RSB for purposes of analytical tractability. For the interested reader and the sake of completeness, we include general results on multiple-step RSB in Appendix A.

4 Limiting Characterization of the Precoder Output

We restrict ourselves in the following to 1RSB analysis of the limiting characteristics of the precoder output. As demonstrated in the sequel, when compared to simulation results at finite numbers of antennas, 1RSB gives quite accurate approximations for the quantities of interest, while the RS ansatz does so only in special cases.

4.1 The 1RSB Solution

Applying the 1RSB ansatz, as outlined in Section 3 (see in particular (3-31)), the limiting properties of the precoder output are characterized by means of four macroscopic parameters $q_1, p_1, \chi_1, \mu_1 \in (0, \infty)$, which are determined as specified below. Let $\boldsymbol{J}$ be a $K \times K$ random matrix satisfying the decomposability property (3-18), and let $R(\cdot)$ denote the R-transform of its limiting eigenvalue distribution. Consider now the following function of complex arguments

$$\beth_u(y,z) \triangleq \mathrm{e}^{-\mu_1 \min\limits_{x\in\mathscr{B}_u} \varepsilon_1|x|^2 - 2\Re\{x(f_1 z^* + g_1 y^*)\}} \quad , \quad (y,z)\in\mathbb{C}^2 \quad , \tag{4-1}$$

where $\Re\{\cdot\}$ takes the real part of the argument, and the parameters ε_1, g_1 and f_1 are defined as

$$\varepsilon_1 = R(-\chi_1) \ , \tag{4-2}$$

$$g_1 = \sqrt{\frac{R(-\chi_1) - R(-\chi_1 - \mu_1 p_1)}{\mu_1}} \ , \tag{4-3}$$

$$f_1 = \sqrt{q_1 R'(-\chi_1 - \mu_1 p_1)} \ . \tag{4-4}$$

Furthermore, denote its normalized version by

$$\tilde{\beth}_u(y,z) = \frac{\beth_u(y,z)}{\int_{\mathbb{C}} \beth_u(\tilde{y},z)\mathrm{d}\tilde{y}} \tag{4-5}$$

to compact notation. Then, using the shortened notations (for $z_{\mathsf{re}}, z_{\mathsf{im}} \in \mathbb{R}$)

$$\int_{\mathbb{C}} (\cdot)\,\mathrm{D}z \triangleq \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (\cdot)\,\frac{\mathrm{e}^{-|z|^2}}{\pi}\,\mathrm{d}z_{\mathsf{re}}\,\mathrm{d}z_{\mathsf{im}} \quad , \quad z \triangleq z_{\mathsf{re}} + j z_{\mathsf{im}} \in \mathbb{C} \quad , \tag{4-6}$$

and

$$\int_{\mathbb{C}^2} (\cdot)\,\mathrm{D}y\,\mathrm{D}z \triangleq \int_{\mathbb{C}}\int_{\mathbb{C}} (\cdot)\,\mathrm{D}y\,\mathrm{D}z, \tag{4-7}$$

the parameters $\{q_1, p_1, \chi_1, \mu_1\}$ are given by the solutions to the four coupled equations[6]

$$\chi_1 + p_1\mu_1 = \frac{1}{f_1}\iint_{\mathbb{C}^2} \Re\left\{ z^* \operatorname*{argmin}_{x\in\mathscr{B}_u} |f_1 z + g_1 y - \varepsilon_1 x| \right\} \tilde{\beth}_u(y,z)\mathrm{D}y\,\mathrm{D}z\,\mathrm{d}F_U(u) \ , \tag{4-8}$$

$$\chi_1 + (q_1 + p_1)\mu_1 = \frac{1}{g_1}\iint_{\mathbb{C}^2} \Re\left\{ y^* \operatorname*{argmin}_{x\in\mathscr{B}_u} |f_1 z + g_1 y - \varepsilon_1 x| \right\} \tilde{\beth}_u(y,z)\mathrm{D}y\,\mathrm{D}z\,\mathrm{d}F_U(u) \ , \tag{4-9}$$

$$q_1 + p_1 = \iint_{\mathbb{C}^2} \left| \operatorname*{argmin}_{x\in\mathscr{B}_u} |f_1 z + g_1 y - \varepsilon_1 x| \right|^2 \tilde{\beth}_u(y,z)\mathrm{D}y\,\mathrm{D}z\,\mathrm{d}F_U(u) \ , \tag{4-10}$$

and

$$\begin{aligned} \int\limits_{\chi_1}^{\chi_1+\mu_1 p_1} R(-w)\,\mathrm{d}w = \iint_{\mathbb{C}} \log\left(\int_{\mathbb{C}} \beth_u(y,z)\,\mathrm{D}y \right) \mathrm{D}z\,\mathrm{d}F_U(u) \\ - 2\chi_1 R(-\chi_1) + (\mu_1 q_1 + 2\chi_1 + 2\mu_1 p_1) R(-\chi_1 - \mu_1 p_1) \\ - 2\mu_1 q_1(\chi_1 + \mu_1 p_1) R'(-\chi_1 - \mu_1 p_1) \ . \end{aligned} \tag{4-11}$$

The limiting properties of the precoder outputs can now be summarized by means of the fol-

[6] In general these coupled equations have multiple solutions and one needs to choose the solution that minimizes the energy penalty.

lowing two propositions. The detailed proofs are provided in Appendices B.1 and B.2, respectively.

Proposition 4.1 *Suppose the random matrix $\boldsymbol{J}$ satisfies the decomposability property* (3-18)*. Then under some technical assumptions, including in particular* one-step replica symmetry breaking*, the effective energy penalty per symbol $\mathscr{E}^{\mathsf{tot}}/K$ converges in probability as $K, N \to \infty$, $K/N \to \alpha < \infty$, to*

$$\bar{\mathscr{E}}_{\mathsf{rsb1}} \triangleq \left(q_1 + p_1 + \frac{\chi_1}{\mu_1}\right) R(-\chi_1 - \mu_1 p_1) - \frac{\chi_1}{\mu_1} R(-\chi_1) - q_1(\chi_1 + \mu_1 p_1) R'(-\chi_1 - \mu_1 p_1) \ . \tag{4-12}$$

The conditional limiting empirical distribution of the precoder's outputs is specified next.

Proposition 4.2 *With the same underlying assumptions as in Proposition 4.1, the limiting conditional empirical distribution of the nonlinear precoder's outputs given an input symbol u satisfies*

$$P_{X|U}(\xi|\upsilon) = \int_{\mathbb{C}^2} 1\left\{\xi = \operatorname*{argmin}_{x \in \mathscr{B}_\upsilon} |f_1 z + g_1 y - \varepsilon_1 x|\right\} \tilde{\daleth}_\upsilon(y, z) \mathrm{D}y \, \mathrm{D}z \ , \tag{4-13}$$

where $1\{\cdot\}$ denotes the indicator function.

4.2 A Replica Symmetric Reduction

Although the 1RSB solution of the replica analysis leads in principle to a more accurate description of the large system limit, corresponding results can also be derived using the simplifying assumption that the system exhibits a replica symmetric behavior (see (3-30)). These results shall be used in the sequel to demonstrate the impact of replica symmetry breaking. However they can also be extremely useful for more conveniently analyzing settings that do exhibit replica symmetric properties, such as the case of convex extended alphabet sets addressed in [25]. A convex alphabet example is discussed in Section 7.

The limiting energy penalty under the RS assumption was in fact already derived in [24], and the result is recalled in the following proposition. The result is given in terms of the two macroscopic parameters $q_0, \chi_0 \in (0, \infty)$, which are obtained through the solution of the two coupled equations

$$q_0 = \int \int_{\mathbb{C}} \left| \operatorname*{argmin}_{x \in \mathscr{B}_u} \left| z - \frac{R(-\chi_0)\, x}{\sqrt{q_0 R'(-\chi_0)}} \right| \right|^2 \mathrm{D}z \, \mathrm{d}F_U(u) \ , \tag{4-14}$$

and

$$\chi_0 = \frac{\int \Re\left\{\int_{\mathbb{C}} \operatorname*{argmin}_{x \in \mathscr{B}_u} \left| z - \frac{R(-\chi_0)\, x}{\sqrt{q_0 R'(-\chi_0)}} \right| z^* \, \mathrm{D}z \right\} \mathrm{d}F_U(u)}{\sqrt{q_0 R'(-\chi_0)}} \ . \tag{4-15}$$

Proposition 4.3 ([24], Proposition 1) *Suppose the random matrix $\boldsymbol{J}$ satisfies the decomposability property* (3-18)*. Then under some technical assumptions, including in particular* replica symmetry*, the effective energy penalty per symbol $\mathscr{E}^{\mathsf{tot}}/K$ converges in probability as $K, N \to \infty$,*

$K/N \to \alpha < \infty$, to[7]

$$\bar{\mathscr{E}}_{rs} \triangleq q_0[R(-\chi_0) - \chi_0 R'(-\chi_0)] \ . \tag{4-16}$$

The limiting conditional distribution of the precoder outputs can also be characterized under the RS assumption, in an analogous manner to Proposition 4.2.

Proposition 4.4 *With the same underlying assumptions as in Proposition 4.3, the limiting conditional empirical distribution of the nonlinear precoder's outputs given an input symbol υ satisfies*

$$P_{X|U}(\xi|\upsilon) = \int_{\mathbb{C}} 1\left\{ \xi = \operatorname*{argmin}_{x \in \mathscr{B}_\upsilon} \left| z - \frac{R(-\chi_0)x}{\sqrt{q_0 R'(-\chi_0)}} \right| \right\} \mathrm{D}z \ . \tag{4-17}$$

This is the measure of the corresponding Voronoi region in the scaled conditional *signal constellation $\mathscr{B}_\upsilon$, with respect to the (complex) Gaussian probability measure.*

Proof: The proof follows the same steps as in the proof of Proposition 4.2, while replacing (3-31) with (3-30). ■

4.3 Zero-Temperature Entropy

One way to demonstrate the degree of consistency of the RS and 1RSB solutions is to look at their limiting (thermodynamic) zero-temperature entropy defined as $\bar{\mathscr{S}} = \lim_{\beta\to\infty} \mathscr{S}(\beta)$. It can also be obtained in a manner similar to Propositions 4.1 and 4.3. In Appendix B.3, we show:

Proposition 4.5 *With the same underlying assumptions as in Propositions 4.1 or 4.3, the limiting entropy per symbol converges to*

$$\bar{\mathscr{S}} = \chi R(-\chi) - \int_0^\chi R(-w)\,\mathrm{d}w \tag{4-18}$$

with χ denoting χ_1 and χ_0 for 1RSB and RS, respectively.

Proposition 4.5 gives rise to the conjecture that the entropy for general r-step RSB is given by $\chi_r R(-\chi_r) - \int_0^{\chi_r} R(-w)\,\mathrm{d}w$ (see (A-1) for the definition of the macroscopic parameters in the general setting).

In any stable thermodynamic system the entropy is non-negative for all temperatures. However, one of the main pitfalls of the RS solution of the original SK-model is that its zero-temperature entropy is negative, indicating an instability [26]. For all R-transforms that are strictly increasing functions of negative real arguments, Proposition 4.5 clearly implies that the entropy is always negative, becoming zero only when the zero temperature value of χ_1, respectively χ_0, approaches zero. While the full RSB solution has been shown to have vanishing entropy at zero temperature and corresponds to the correct solution, the following lemma proven in Appendix E, indicates that negative entropy is a rather common effect for finite RSB steps.

[7] In [24], the self-averaging property was stated as an assumption, since the authors were not aware of [34].

Lemma 4.6 *The R-transform, wherever its derivative with respect to a real argument exists, is an increasing function. If the probability distribution is different from a single mass point, the R-transform is strictly increasing.*

Note that the above argument for the entropy holds only for discrete state variables. In the case of continuous alphabets, the (then differential) entropy of a system can in fact be negative. Therefore, a negative zero-temperature entropy is not an alarm bell *per se*. For discrete state variables, the zero-temperature entropy serves as a measure of accuracy: the closer it is to zero, the better the approximation.

5 Zero-Forcing Front-End

To gain more insight into the impact on system performance of the nonlinear precoding scheme under investigation, we now particularize to a specific linear front-end, namely the ZF front-end. The precoding matrix $\boldsymbol{T}$ in this case is given by the pseudo-inverse of the channel transfer matrix, which we write here as

$$\boldsymbol{T} = \boldsymbol{H}^{+} = \lim_{\epsilon \to 0} \boldsymbol{H}^{\dagger} \left(\boldsymbol{H}\boldsymbol{H}^{\dagger} + \epsilon \mathbf{I} \right)^{-1} \quad . \tag{5-1}$$

The underlying assumptions are that $N \geq K$ and that the matrix $\boldsymbol{H}\boldsymbol{H}^{\dagger}$ is almost surely (a.s.) positive definite[8]. Focusing on the asymptotic regime for $K/N \to \alpha \leq 1$, then using (2-1), (2-3), and Proposition 4.1, the equivalent single-user channel observed by user i is

$$\check{r}_i \approx x_i + \check{n}_i, \quad K \gg 1 \ , \tag{5-2}$$

where $\check{n}_i$ is a zero mean circularly symmetric complex Gaussian noise with variance $\frac{1}{\rho}$,

$$\rho \triangleq \frac{\mathsf{snr}}{\bar{\mathscr{E}}_{\mathsf{rsb1}}} \tag{5-3}$$

denotes the effective received SNR, and $\bar{\mathscr{E}}_{\mathsf{rsb1}}$ is given by (4-12).

Proposition 5.1 *Employing the same underlying assumptions as in Proposition 4.1, then with a ZF front-end the channel observed by a randomly chosen user is equivalent in the large system limit to a concatenated single-user channel, with input $u \in \mathscr{U}$, intermediate output $x \in \mathscr{B}_u$, and final output $y \in \mathbb{C}$, specified by the Markov chain u–x–y as shown in Figure 2. This Markov chain is defined by the following joint probability density function*

$$f_{UXY}(u, x, y) = f_U(u) f_{X|U}(x|u) f_{Y|X}(y|x) \quad , \tag{5-4}$$

[8] In Section 8, we will also allow for $N < K$ following the treatment in [42].

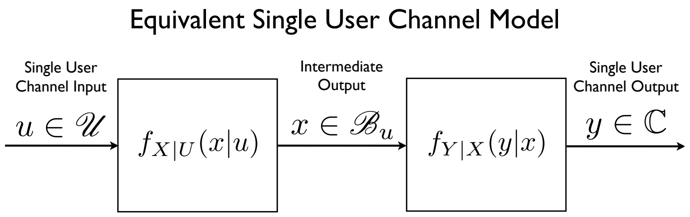

Figure 2: Schematic description of the equivalent single user channel model for a ZF front-end.

where

$$f_{X|U}(x|u) = \sum_{\tilde{x}\in\mathscr{B}_u} P_{X|U}(\tilde{x}|u)\delta(x-\tilde{x}) \quad , \tag{5-5}$$

with $P_{X|U}(\mathsf{x}|u)$ *given by* (4-13)*, and*

$$f_{Y|X}(y|x) = \frac{\rho}{\pi}\mathrm{e}^{\,-|y-x|^2\rho} \tag{5-6}$$

is the (complex) Gaussian density with mean x *and variance* $1/\rho$*.*

Proof: The Proposition follows straightforwardly from Proposition 4.2 and (5-2). ■

Note that the RS reduction of the above result is readily obtained from Propositions 4.3 and 4.4, by replacing $\bar{\mathscr{E}}_{\mathsf{rsb1}}$ in (5-3) with $\bar{\mathscr{E}}_{\mathsf{rs}}$ of (4-16), and taking (4-17) for $P_{X|U}(\tilde{x}|u)$ in (5-5).

The achievable throughput of the nonlinear precoding scheme can be derived from the equivalent single-user channel model using Proposition 5.1. Accordingly, the achievable rate of a randomly chosen user is given by the mutual information[9] between the input u and received signal y, i.e.,

$$R = \mathrm{I}(u\,;y) = \mathrm{h}(y) - \mathrm{h}(y|u) \quad , \tag{5-7}$$

where $\mathrm{h}(\cdot)$ and $\mathrm{h}(\cdot|\cdot)$ denote differential entropy and conditional differential entropy, respectively (which can be readily calculated using Proposition 5.1). The *normalized* spectral efficiency is then given by

$$C \approx \frac{K}{N}R \xrightarrow[K\to\infty]{} \alpha R \quad , \tag{5-8}$$

and it is functionally dependent on the system average E_b/N_0 through the relation [43]

$$\mathsf{snr} = \frac{1}{\alpha}C\frac{E_b}{N_0} \quad . \tag{5-9}$$

To get a better insight into the impact of the nonlinear precoding scheme, it is useful to compare the results to the spectral efficiency of DPC with Gaussian input (specifying the ultimate performance), as well as to the spectral efficiency of *linear* ZF (for both Gaussian and discrete

[9] Note that in the large-system limit the receivers only need information about the state of their own channel, but not about the states of the other channels due to the self-averaging property which makes the impact of the other users' channels and data deterministic.

alphabet input). Another interesting comparison is to the spectral efficiency of generalized THP (GTHP), which is a popular practical nonlinear precoding alternative to the scheme considered here (see, e.g., [20]). For the sake of comparison we further particularize henceforth to the case in which the entries of the channel transfer matrix $\boldsymbol{H}$ are i.i.d. zero-mean circularly symmetric complex Gaussian random variables, with variance $1/N$ ("a Gaussian $\boldsymbol{H}$"). Note that in this case the R-transform of the limiting eigenvalue distribution of the random matrix $\boldsymbol{J} = (\boldsymbol{H}\boldsymbol{H}^\dagger)^{-1}$, and its derivative, simplifiy to [24]

$$R(w) = \frac{1-\alpha-\sqrt{(1-\alpha)^2-4\alpha w}}{2\alpha w} \quad , \tag{5-10}$$

$$R'(w) = \frac{\left(1-\alpha-\sqrt{(1-\alpha)^2-4\alpha w}\right)^2}{4\alpha w^2\sqrt{(1-\alpha)^2-4\alpha w}} \quad . \tag{5-11}$$

Starting with DPC, the limiting spectral efficiency in this setting coincides with the corresponding spectral efficiency of the *dual uplink* channel with uniform power distribution [43]. This follows from the limiting conclusion in [44], and by observing that the optimization problem over diagonal input covariance matrices, that specifies the maximum achievable sum-rate (see [1] and references therein), is solved by a uniform power distribution [45]. The spectral efficiency of DPC is hence given by [43]

$$C^{\mathsf{dpc}}(\mathsf{snr}) = \alpha \log_2\left(1+\mathsf{snr}-\frac{1}{4}\mathcal{F}(\mathsf{snr},\alpha)\right) + \log_2\left(1+\alpha\,\mathsf{snr}-\frac{1}{4}\mathcal{F}(\mathsf{snr},\alpha)\right) - \frac{\log_2 \mathrm{e}}{4\mathsf{snr}}\mathcal{F}(\mathsf{snr},\alpha) \ , \tag{5-12}$$

where $\mathcal{F}(\mathsf{snr},\alpha)$ is defined as [46]

$$\mathcal{F}(\mathsf{snr},\alpha) \triangleq \left(\sqrt{\mathsf{snr}\,(1+\sqrt{\alpha})^2+1}-\sqrt{\mathsf{snr}\,(1-\sqrt{\alpha})^2+1}\right)^2 \quad . \tag{5-13}$$

Regarding linear ZF, we restrict the discussion to the case in which the active user population can only be controlled through the system load α, as is in fact assumed for the nonlinear precoding scheme (see also the discussion in Section 8). In this setting, as shown, e.g., in [47,48], the induced precoding efficiency (2-7) (equivalent here to the inverse multiuser efficiency) converges in the large system limit to

$$\zeta_{\mathsf{zf}} = \frac{1}{1-\alpha} \quad , \tag{5-14}$$

and again for Gaussian input the spectral efficiency coincides with the corresponding result in [48] (see also [43, 46, 49])

$$C^{\mathsf{zf}}(\mathsf{snr}) = \alpha \log_2\left(1+(1-\alpha)\,\mathsf{snr}\right) \quad . \tag{5-15}$$

The corresponding spectral efficiency with discrete input alphabet can be derived, e.g., following the guidelines in [50]. Considering the particular case of binary phase shift keying (BPSK) input, one obtains

$$C^{\mathsf{zf,bpsk}}(\mathsf{snr}) = \alpha \left(1 - \int_{-\infty}^{\infty} \sqrt{\frac{(1-\alpha)\,\mathsf{snr}}{\pi}} \mathrm{e}^{\,-(1-\alpha)\,\mathsf{snr}\,(s-1)^2} \log_2 \left(1 + \mathrm{e}^{\,-4(1-\alpha)\,\mathsf{snr}\,s} \right) \mathrm{d}s \right) \quad . \tag{5-16}$$

The spectral efficiency of linear ZF precoding combined with QPSK input is obtained via the relation [50]

$$C^{\mathsf{zf,qpsk}}(\mathsf{snr}) = 2C^{\mathsf{zf,bpsk}}\left(\frac{\mathsf{snr}}{2}\right) \quad , \tag{5-17}$$

yielding (through (5-9)) $C^{\mathsf{zf,qpsk}}(\frac{E_b}{N_0}) = 2C^{\mathsf{zf,bpsk}}\left(\frac{E_b}{N_0}\right)$. The spectral efficiency of GTHP for the corresponding setting is derived in Appendix F.

6 Lattice Precoding: An RSB Example

Adhering to [24], we consider in the following a particular example of a discrete relaxed alphabet set for QPSK signaling, which exhibits replica symmetry breaking. The original QPSK constellation alphabet is represented by the set

$$\mathscr{U} = \{1+j, -1+j, -1-j, 1-j\} \quad , \tag{6-1}$$

and quadrature symmetric transmissions are assumed (note that the above definition induces $\sigma_u^2 = 2$). The relaxed alphabets in this particular example can be represented as points from the extended lattice

$$\mathscr{B}_u = \frac{u}{1+j}((4\mathbb{Z}+1) \times (4\mathbb{Z}+1)) \quad , \quad \forall u \in \mathscr{U} \quad . \tag{6-2}$$

More specifically, we take

$$\mathscr{B}_{\pm 1 \pm j} = \pm \{c_1, c_2, \dots, c_L\} \pm j\, \{c_1, c_2, \dots, c_L\} \quad , \tag{6-3}$$

where it is assumed that $-\infty = c_0 < c_1 < \cdots < c_L < c_{L+1} = \infty$. The parameter L thus specifies the number of lattice points used in the extended alphabet in each dimension, and we particularize here to the set $\{+1, -3, +5, -7, +9, \dots\}$. The alphabet relaxation scheme is depicted in Figure 3. Due to the complete quadrature symmetry of this setting, all QPSK constellation points and their corresponding relaxed alphabet subsets are completely equivalent, and we focus in the following, for notational convenience, on the QPSK constellation point represented by $u = 1+j$, and $\mathscr{B}_{1+j}$.

The first step in the analysis is to rewrite (4-8)–(4-11) and obtain the four macroscopic parameters $\{q_1, p_1, \chi_1, \mu_1\}$ for the current example. Denoting the real and imaginary parts of an arbitrary point $s \in \mathbb{C}$ as $s_{\mathsf{re}} \triangleq \Re\{s\}$ and $s_{\mathsf{im}} \triangleq \Im\{s\}$, the Voronoi region of the lattice point $x = c_m + jc_n$ is the region in the complex plane for which

$$s_{\mathsf{re}} \in (v_m, v_{m+1}) \quad , \quad s_{\mathsf{im}} \in (v_n, v_{n+1}) \quad , \tag{6-4}$$

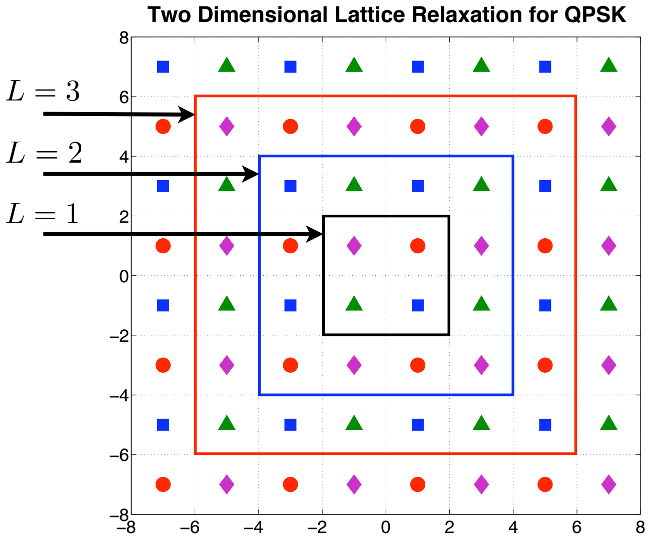

Figure 3: Two dimensional lattice based relaxation for QPSK input.

where the boundaries of the Voronoi regions are $\left\{v_i = \frac{c_i+c_{i-1}}{2}\right\}$. Now considering (4-1), recall that for any given $y, z \in \mathbb{C}$ the lattice point that maximizes the exponent therein is given by $\underset{x\in\mathscr{B}_u}{\operatorname{argmin}} |f_1 z + g_1 y - \varepsilon_1 x|$. This implies that a lattice point $x = c_m + jc_n$ is the solution to the above minimization problem whenever

$$y_{\mathsf{re}} \in (\psi_m(z_{\mathsf{re}}), \psi_{m+1}(z_{\mathsf{re}})) \ , \quad y_{\mathsf{im}} \in (\psi_n(z_{\mathsf{im}}), \psi_{n+1}(z_{\mathsf{im}})) \ , \tag{6-5}$$

where we introduced the real argument function

$$\psi_k(\xi) \triangleq \frac{\varepsilon_1 v_k - f_1 \xi}{g_1} \quad . \tag{6-6}$$

Applying this observation to (4-8)–(4-11), and exploiting the quadrature symmetry property, the derivation simplifies considerably by noticing that the inner integrals therein can be represented as sums of separate integrals over the regions specified by (6-5). Accordingly, consider the two real argument functions

$$\Theta_k(\xi) \triangleq \mathrm{e}^{\mu_1 c_k\left[(\mu_1 g_1^2 - \varepsilon_1)c_k + 2f_1\xi\right]}\left[Q\left(\sqrt{2}(\psi_k(\xi) - \mu_1 g_1 c_k)\right) - Q\left(\sqrt{2}(\psi_{k+1}(\xi) - \mu_1 g_1 c_k)\right)\right] \ , \tag{6-7}$$

$$\Psi_k(\xi) \triangleq \frac{1}{2\sqrt{\pi}}\mathrm{e}^{\mu_1 c_k\left[(\mu_1 g_1^2 - \varepsilon_1)c_k + 2f_1\xi\right]}\left[\mathrm{e}^{-(\psi_k(\xi) - \mu_1 g_1 c_k))^2} - \mathrm{e}^{-(\psi_{k+1}(\xi) - \mu_1 g_1 c_k))^2}\right] \ . \tag{6-8}$$

Then, following some tedious algebra, it can be shown from (4-8)–(4-11) that the parameters

$\{q_1, p_1, \chi_1, \mu_1\}$ are the solutions to the coupled equations

$$q_1 = 2\int_{-\infty}^{\infty} \frac{\sum_{m=1}^{L} c_m^2 \Theta_m(\xi)}{\sum_{k=1}^{L} \Theta_k(\xi)} \mathrm{e}^{-\xi^2} \frac{\mathrm{d}\xi}{\sqrt{\pi}} - p_1 \quad , \tag{6-9}$$

$$p_1 = \frac{2}{f_1 \mu_1} \int_{-\infty}^{\infty} \frac{\sum_{m=1}^{L} c_m \Theta_m(\xi)}{\sum_{k=1}^{L} \Theta_k(\xi)} \xi \mathrm{e}^{-\xi^2} \frac{\mathrm{d}\xi}{\sqrt{\pi}} - \frac{\chi_1}{\mu_1} \ , \tag{6-10}$$

$$\chi_1 = \frac{2}{g_1} \int_{-\infty}^{\infty} \frac{\sum_{m=1}^{L} c_m \Psi_m(\xi)}{\sum_{k=1}^{L} \Theta_k(\xi)} \mathrm{e}^{-\xi^2} \frac{\mathrm{d}\xi}{\sqrt{\pi}} \ , \tag{6-11}$$

and

$$\mu_1 = \left[2q_1(\chi_1 + \mu_1 p_1) R'(-\chi_1 - \mu_1 p_1)\right]^{-1} \cdot \left[2\int_{-\infty}^{\infty} \log\left(\sum_{m=1}^{L} \Theta_m(\xi)\right) \mathrm{e}^{-\xi^2} \frac{\mathrm{d}\xi}{\sqrt{\pi}} \right.$$
$$\left. - \int_{\chi_1}^{\chi_1 + \mu_1 p_1} R(-w)\, \mathrm{d}w - 2\mu_1 \chi_1 g_1^2 + \mu_1 (q_1 + 2p_1) R(-\chi_1 - \mu_1 p_1) \right] . \tag{6-12}$$

The corresponding energy penalty is obtained by plugging the four solutions into (4-12). Applying the same approach to Proposition 4.2, the limiting conditional probability of the precoder output being $x = c_m + jc_n \in \mathscr{B}_{1+j}$ is given by

$$\Pr\{x = c_m + jc_n \in \mathscr{B}_u | u = 1+j\} = \int_{-\infty}^{\infty} \frac{\Theta_m(\xi)}{\sum_{k=1}^{L} \Theta_k(\xi)} \mathrm{e}^{-\xi^2} \frac{\mathrm{d}\xi}{\sqrt{\pi}} \cdot \int_{-\infty}^{\infty} \frac{\Theta_n(\zeta)}{\sum_{\ell=1}^{L} \Theta_\ell(\zeta)} \mathrm{e}^{-\zeta^2} \frac{\mathrm{d}\zeta}{\sqrt{\pi}}$$
$$= \Pr\{\Re\{x\} = c_m | u = 1+j\} \cdot \Pr\{\Im\{x\} = c_n | u = 1+j\} \ . \tag{6-13}$$

The limiting conditional probabilities that correspond to the rest of the QPSK constellation points are readily obtained from (6-13) by symmetry considerations. Note also that (6-13) implies that the real part and the imaginary part of the precoder output x behave as independent random variables.

Numerical results for the limiting energy penalty of the discrete lattice relaxation scheme are plotted in Figure 4. The figure shows the limiting energy penalty (in dB) as a function of the system load α, for the particular case of a Gaussian $\boldsymbol{H}$ and a ZF front-end. Since $\sigma_u^2 = 2$, the corresponding precoding efficiency (2-7) can be immediately obtained by subtracting 3dB from the energy penalty shown in the figure. The results in Figure 4 correspond to alphabet relaxations with $L = 2$ and $L = 3$. Note that the two curves are essentially indistinguishable and the energy penalty with $L = 2$ becomes only negligibly larger as α gets close to unity. This implies that increasing L beyond 2 in this setting provides diminishing returns. Empirical energy penalties obtained through Monte Carlo simulations are also included in the figure. The results are for systems in which the *number of users* is fixed to $K = 8$, $K = 16$, and $K = 32$ (averaged over 10^4, 10^3, and 10^2 channel realizations, respectively). The energy penalty is shown to decrease with the system size, and the simulation results exhibit a good match to the limiting energy penalty predicted by the 1RSB replica analysis. The lower bound for the limiting energy penalty obtained in [21] is also plotted in this figure which, with appropriate scaling to match the current setting,

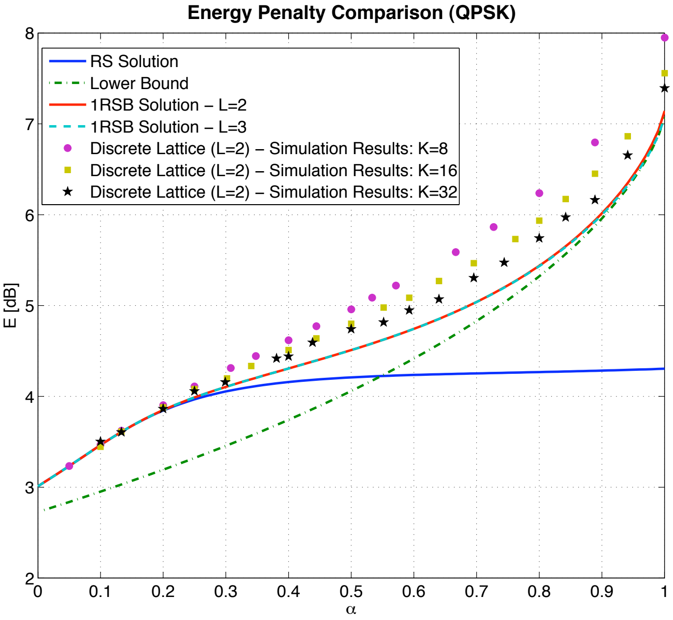

Figure 4: The energy penalty per symbol, as a function of the system load α, for the two dimensional discrete alphabet relaxation scheme for QPSK input.

is given by

$$\bar{\mathscr{E}}_{LB} = \frac{16}{\pi}(1-\alpha)^{\frac{1}{\alpha}-1} \quad . \tag{6-14}$$

Figure 4 shows that the 1RSB prediction approaches the lower bound as the load approaches unity. Note however that the 1RSB result stays strictly higher than the lower bound. In fact, a careful numerical examination of the limiting 1RSB energy penalty at $\alpha = 1$ shows that it hits the value of 7.0744 dB for $L \geq 4$, while the lower bound in this case is $\frac{16}{\pi} \approx 7.0697$ dB. The numerical analysis of the limiting energy penalty is considerably simplified in this region of α by the (numerical) observation that the macroscopic parameter χ_1 approaches 0 as $\alpha \to 1$ (although it stays strictly positive). The small χ_1 approximation of the equations employed to calculate the limiting 1RSB energy penalty is shortly discussed in Appendix D. The RSB phenomena is demonstrated by considering the limiting energy penalty obtained via the RS approximation, as stated by Proposition 4.3 (the explicit expression for the current example is given in [24, Eq. (26)]). As shown in Figure 4, the RS approximation fails to predict the limiting energy penalty for $\alpha > 0.3$, and in fact it even violates the lower bound (6-14) for $\alpha > 0.55$.

The better accuracy of 1RSB is also visible looking at the zero-temperature entropy. We can analytically evaluate Proposition 4.5 in the case of a Gaussian $\boldsymbol{H}$, which becomes

$$\bar{\mathscr{S}} = \frac{1-\alpha-\sqrt{(1-\alpha)^2+4\alpha\chi}}{2\alpha} + \frac{1-\alpha}{\alpha}\log\left(\frac{1-\alpha+\sqrt{(1-\alpha)^2+4\alpha\chi}}{2(1-\alpha)}\right) \quad . \tag{6-15}$$

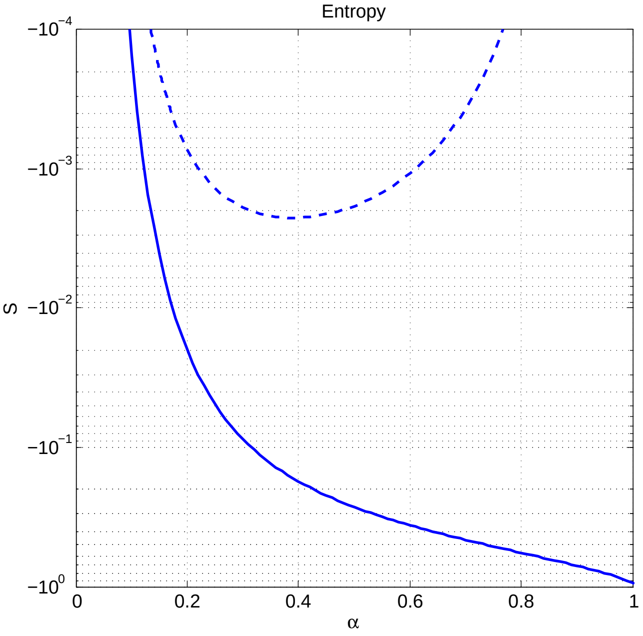

Figure 5: Zero temperature entropy of the RS solution (solid), and 1RSB solution (dashed), as a function of the system load α (corresponding to (6-15)).

The entropy for both the RS and 1RSB approximations for a relaxation level $L = 2$ are shown in Figure 5. Although the 1RSB solution of the above model also has negative zero-temperature entropy, it is much closer to zero, corresponding to a much weaker instability, and approaches zero as $\alpha \to 1$. In contrast, the RS entropy drifts away from zero as $\alpha \to 1$.

The limiting conditional probabilities of (6-13) are plotted in Figure 6, as well as the empirical conditional probabilities, based on the Monte Carlo simulations employed to produce the energy penalties of Figure 4. The results correspond to a relaxation level of $L = 2$, and focus on the *real part* of the extended alphabet points, given that the real part of the original QPSK constellation point satisfies $\Re\{u\} = 1$ (recall the decoupling of the real and imaginary parts implied by (6-13)). The simulation results exhibit again a good match to the limiting analytical 1RSB prediction. It is also clearly demonstrated that, when the system load α is low, hardly any relaxation is required, while the probability of using symbols from the extended alphabet set increases as α approaches unity.

7 Convex Precoding: An RS Example

This section is devoted to another alphabet relaxation scheme, also introduced in [24] for QPSK signaling. The key feature of this relaxation scheme is that the extended alphabet set is continuous and *convex*, allowing for an efficient solution to the corresponding quadratic programming problem of minimizing the energy penalty. Convex optimization problems are generally believed not to

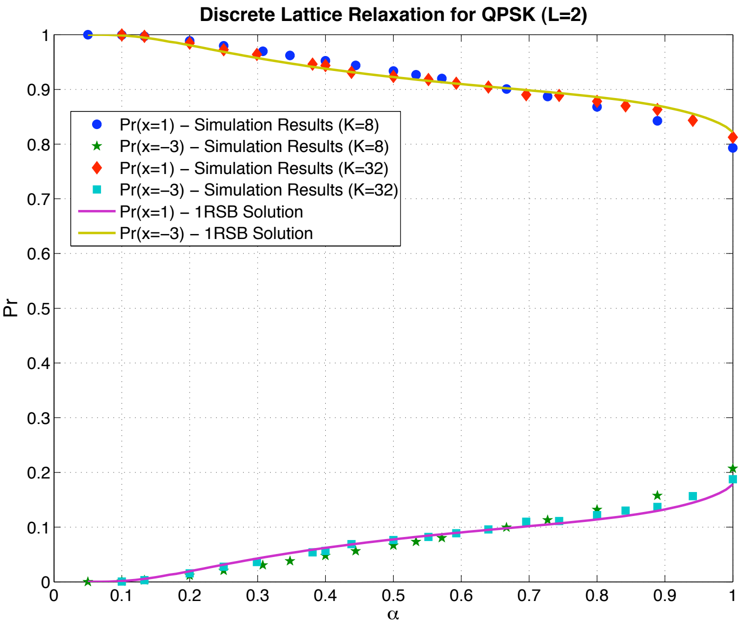

Figure 6: Conditional probabilities of the real part of the precoder output for the two-dimensional discrete lattice relaxation scheme, given that $\Re\{u\} = 1$.

exhibit replica symmetry breaking [51]. In certain special cases this has been shown explicitly [52]. Furthermore, as will be demonstrated in the sequel, the *replica symmetric* solution for this alphabet relaxation scheme agrees well with numerical simulations and thus considerably simplifies the analysis of the limiting regime.

Denoting

$$\mathscr{B}_{1+j} = \{z \in \mathbb{C} : \Re\{z\} \geq 1, \Im\{z\} \geq 1\} \quad , \tag{7-1}$$

the relaxed alphabet subsets are defined by

$$\mathscr{B}_u = \frac{u}{1+j}\mathscr{B}_{1+j} \quad , \quad u \in \{1+j, -1+j, -1-j, 1-j\} \quad . \tag{7-2}$$

The alphabet relaxation scheme is depicted in Figure 7, and it is referred to henceforth as *convex relaxation for QPSK (CR-QPSK)*.

The RS approximation for the limiting energy penalty with the CR-QPSK relaxation scheme is obtained through Proposition 4.3, and it is given by the solution to the following fixed point equation [24, Eq. (30)]

$$Q\left(\sqrt{\frac{2}{\alpha\bar{\mathscr{E}}}}\right) = \frac{2 + (\alpha-1)\bar{\mathscr{E}} + \sqrt{\frac{\alpha\bar{\mathscr{E}}}{\pi}}\mathrm{e}^{-\frac{1}{\alpha\bar{\mathscr{E}}}}}{2 + \alpha\bar{\mathscr{E}}} \quad . \tag{7-3}$$

Note that (7-3) yields finite energy penalties for all loads $0 \leq \alpha < 2$. Although loads greater than

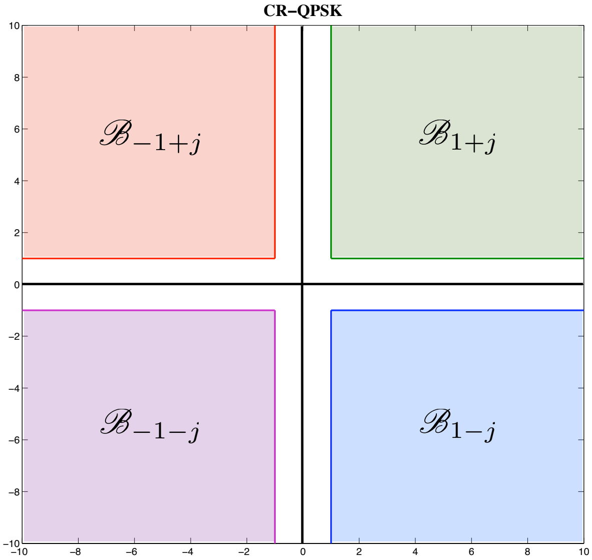

Figure 7: Extended alphabet sets for the convex relaxation precoding scheme for QPSK signaling.

unity imply that the matrix $\boldsymbol{H}\boldsymbol{H}^\dagger$ in (5-1) is singular, this does not lead to interference at the receivers in the large system limit, as shown rigorously in [42].

Numerical results for the limiting energy penalty of CR-QPSK are plotted in Figure 8. Empirical results based on Monte Carlo simulations are provided as well. These results were obtained by fixing the number of users to $K = 32$, and averaging over 1000 channel realizations. The results exhibit an excellent match to the limiting RS analytical results, thus supporting the validity of the RS approximation. The corresponding results for the discrete lattice-based alphabet relaxation scheme of Section 6 are also provided for the sake of comparison, and it is clearly observed that in terms of the limiting energy penalty, the discrete scheme is superior to the CR-QPSK scheme for all $\alpha \in (0, 1]$. The limiting energy penalty difference approaches its maximum value of 2.41 dB at $\alpha = 1$. As will be shown in Section 8, however, the comparison becomes more subtle when spectral efficiency is investigated.

The RS approximation of the limiting conditional distribution of the precoder outputs is obtained using Proposition 4.4. The idea here is to start from a discretized version of the continuous CR-QPSK relaxed alphabet set, and obtain the limiting conditional distribution of each relaxed alphabet point using (4-17). The final step is then to take the limit as the areas of the Voronoi cells corresponding to each such point vanish. Using this approach, while restricting the discussion to a Gaussian $\boldsymbol{H}$ and focusing for convenience on the QPSK constellation point $u = 1 + j$, one

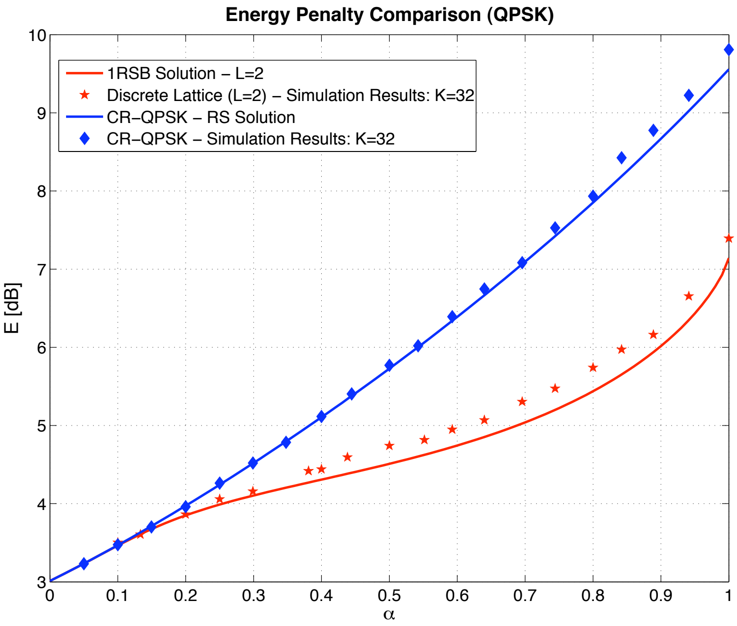

Figure 8: The energy penalty per symbol as a function of the system load α, for the CR-QPSK alphabet relaxation scheme. Corresponding results for the discrete alphabet relaxation scheme of Section 6 are provided for comparison.

gets the corresponding conditional probability *density* function (pdf)

$$\begin{aligned} f_{X|U}^{\textsf{cr-qpsk}}(x|u=1+j) = Q_1^2\delta(x_{\textsf{re}}-1)\delta(x_{\textsf{im}}-1) \\ + Q_1\frac{1}{\sqrt{\pi\alpha\bar{\mathscr{E}}^{\textsf{cr-qpsk}}}}\mathrm{e}^{-\frac{x_{\textsf{im}}^2}{\alpha\bar{\mathscr{E}}^{\textsf{cr-qpsk}}}}\delta(x_{\textsf{re}}-1)\mathcal{U}(x_{\textsf{im}}-1) \\ + Q_1\frac{1}{\sqrt{\pi\alpha\bar{\mathscr{E}}^{\textsf{cr-qpsk}}}}\mathrm{e}^{-\frac{x_{\textsf{re}}^2}{\alpha\bar{\mathscr{E}}^{\textsf{cr-qpsk}}}}\delta(x_{\textsf{im}}-1)\mathcal{U}(x_{\textsf{re}}-1) \\ + \frac{1}{\pi\alpha\bar{\mathscr{E}}^{\textsf{cr-qpsk}}}\mathrm{e}^{-\frac{|x|^2}{\alpha\bar{\mathscr{E}}^{\textsf{cr-qpsk}}}}\mathcal{U}(x_{\textsf{re}}-1)\mathcal{U}(x_{\textsf{im}}-1)\ , \quad x_{\textsf{re}},x_{\textsf{im}}\in\mathbb{R} \quad , \end{aligned} \tag{7-4}$$

where we decompose the complex argument as $x \triangleq x_{\textsf{re}}+jx_{\textsf{im}}$, $\mathcal{U}(x)$ denotes the unit step function, $\bar{\mathscr{E}}^{\textsf{cr-qpsk}}$ denotes the limiting energy penalty of the CR-QPSK scheme obtained from (7-3), and the constant Q_1 is defined as

$$Q_1 \triangleq Q\left(-\sqrt{\frac{2}{\alpha\bar{\mathscr{E}}^{\textsf{cr-qpsk}}}}\right) \quad . \tag{7-5}$$

The conditional pdf given the rest of the QPSK constellation points (i.e., $u\in\{-1+j,\ -1-j, 1-j\}$) is obtained in an analogous manner, while considering the full symmetry of the extended constellation.

Returning to (7-4), note that this pdf contains masses on the boundaries of $\mathscr{B}_{1+j}$, and in particular a mass point at the original QPSK constellation point (i.e., $x=1+j$). Plots that

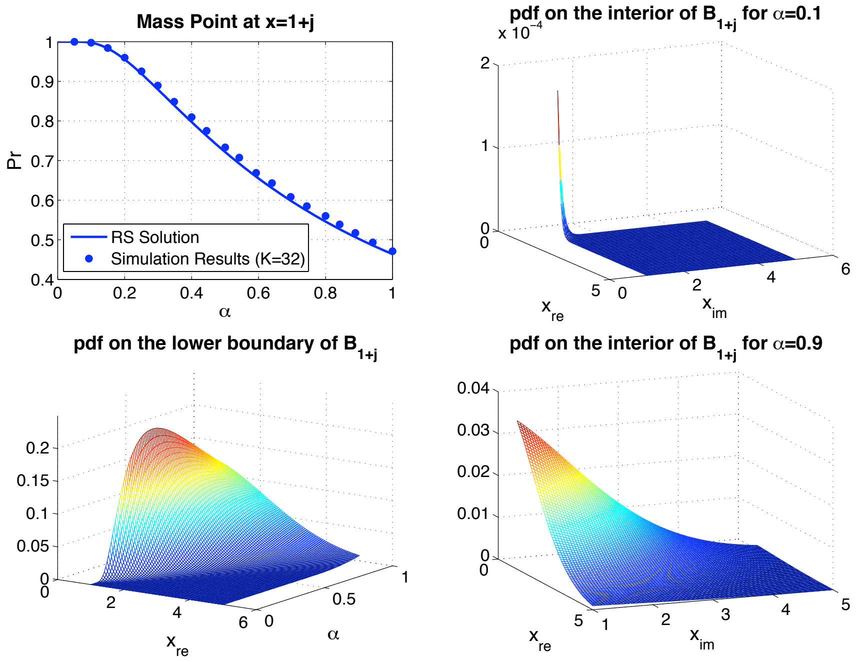

Figure 9: The limiting conditional distribution of the precoder output for the CR-QPSK scheme, given that $u = 1 + j$.

demonstrate this behavior of the pdf as a function of α are provided in Figure 9. The upper left plot shows the weight of the mass point at $x = 1+j$, as a function of α (corresponding to Q_1^2). The lower left plot shows the pdf mass on the lower boundary of the extended alphabet subset $\mathscr{B}_{1+j}$ (i.e., when the imaginary part of the precoder's output is fixed to $x_{\mathsf{im}} = j$). The plots on the right show the pdf on the interior of $\mathscr{B}_{1+j}$, for $\alpha = 0.1$ (upper right) and $\alpha = 0.9$ (lower right). The increase in probability of using extended alphabet points as the system load increases, is clearly demonstrated in the figure.

Additional numerical results comparing the analytical RS approximation for the pdf to empirical simulation results are shown in the upper left plot of Figure 9 and in Figure 10. The upper left plot of Figure 9 compares the probability mass at $x = 1 + j$ to corresponding simulation results for $K = 32$ (averaged over 1000 channel realizations). The corresponding comparison for the cumulative distribution function (CDF) of x_{re}, given that $\Re\{u\} = 1$, is shown in Figure 10. The left plot shows the CDF for $\alpha = 0.7442$ (i.e., for $N = 43$), while the right plot shows the results for unit load. As observed, all empirical results exhibit a very good match to the analytical RS approximation, further supporting the validity of the RS analysis for the CR-QPSK scheme.

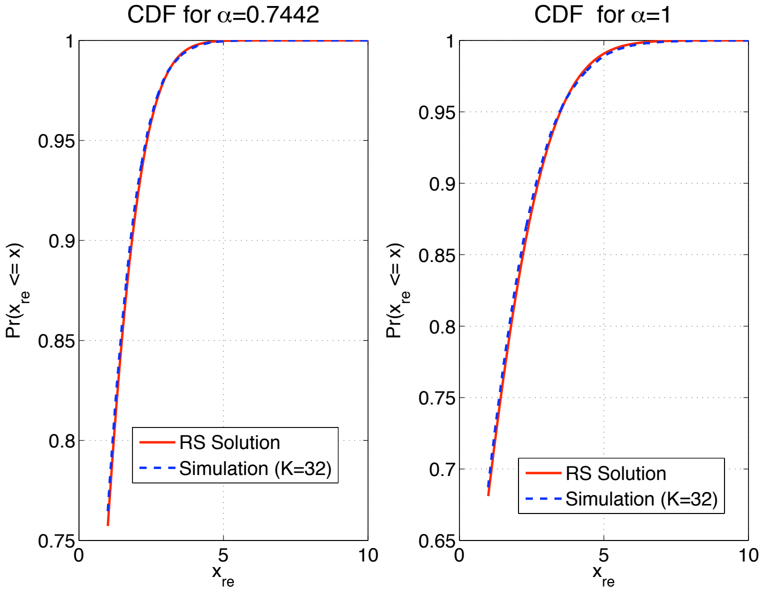

Figure 10: CDF of the *real* part of the precoder's output for the CR-QPSK scheme, given that $\Re\{u\} = 1$.

8 Spectral Efficiency Comparison

The two previous sections focused on the *transmitting* end of the system, and investigated the limiting behavior of the precoder output while employing two particular alphabet relaxation schemes. In the following we turn to investigate the limiting behavior of the system as a whole, by considering the normalized spectral efficiency in view of the analysis of Section 5. Accordingly, we restrict the discussion to a ZF front-end and a Gaussian $\boldsymbol{H}$, and apply Proposition 5.1 to obtain the spectral efficiencies of the discrete lattice-based alphabet relaxation scheme and of CR-QPSK.

Starting with the discrete scheme, the spectral efficiency is obtained by incorporating (4-12) and (6-13) into (5-4)–(5-8). The observation made in Section 6 regarding the independence of the real and imaginary parts of the precoder's output, leads to the following conclusion. The achievable rate in (5-7) for QPSK input can be obtained by treating QPSK signaling as two independent corresponding BPSK signaling settings. Accordingly, the conditional precoder output probabilities, given a real BPSK input of $u = 1$, are given by (cf. (6-13))

$$P^{\mathsf{bpsk}}_{X|u=1}(c_k) \quad \triangleq \quad \int_{-\infty}^{\infty} \frac{\Theta_k(\xi)}{\sum_{m=1}^{L} \Theta_m(\xi)}\, \mathrm{e}^{-\xi^2} \frac{\mathrm{d}\xi}{\sqrt{\pi}} \ , \tag{8-1}$$

where we set $\mathscr{B}_1 = \{c_k\}_{k=1}^{L}$, and the conditional probabilities given $u = -1$ can be immediately obtained from symmetry considerations. It is then straightforward to show that the corresponding spectral efficiency is given by

$$C^{\textsf{bpsk}}(\rho) = \alpha\left[1 - \int_{-\infty}^{\infty} \sum_{k=1}^{L} P^{\textsf{bpsk}}_{X|u=1}(c_k)\sqrt{\frac{\rho}{\pi}}\mathrm{e}^{\,-\rho(\xi-c_k)^2}\log_2\left(1+\frac{\sum_{k=1}^{L} P^{\textsf{bpsk}}_{X|u=1}(c_k)\mathrm{e}^{\,-\rho(\xi+c_k)^2}}{\sum_{k=1}^{L} P^{\textsf{bpsk}}_{X|u=1}(c_k)\mathrm{e}^{\,-\rho(\xi-c_k)^2}}\right)\mathrm{d}\xi\right]. \tag{8-2}$$

The spectral efficiency with QPSK input is then obtained through the relation $C^{\textsf{qpsk}}(\frac{E_b}{N_0}) = 2C^{\textsf{bpsk}}(\frac{E_b}{N_0})$, while substituting

$$\rho = \frac{C\frac{E_b}{N_0}}{\alpha\bar{\mathcal{E}}_{\textsf{rsb1}}} \quad . \tag{8-3}$$

Turning to the CR-QPSK scheme, and applying Proposition 5.1, it can be shown that the conditional pdf of the equivalent single user channel output y, given an input u, is equal to[10]

$$\begin{aligned} f^{\textsf{cr-qpsk}}_{Y|U}(y|u=\pm1\pm j) = &\sqrt{\frac{\rho}{\pi}}\left[\frac{1}{\sqrt{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}}Q_2(\pm y_{\textsf{re}})\mathrm{e}^{\,-\frac{y_{\textsf{re}}^2\rho}{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}} + Q_1\mathrm{e}^{\,-(y_{\textsf{re}}\mp1)^2\rho}\right] \\ &\cdot\sqrt{\frac{\rho}{\pi}}\left[\frac{1}{\sqrt{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}}Q_2(\pm y_{\textsf{im}})\mathrm{e}^{\,-\frac{y_{\textsf{im}}^2\rho}{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}} + Q_1\mathrm{e}^{\,-(y_{\textsf{im}}\mp1)^2\rho}\right] \quad , \end{aligned} \tag{8-4}$$

where we decomposed the complex argument as $y \triangleq y_{\textsf{re}} + jy_{\textsf{im}}$, and the real argument function $Q_2(\xi)$ is defined as

$$Q_2(\xi) \triangleq Q\left(\sqrt{\frac{2}{\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}}\frac{\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}(1-\xi)+1}{\sqrt{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}}\right) \quad , \quad \xi\in\mathbb{R} \quad . \tag{8-5}$$

The marginal distribution of the equivalent single user channel output y is given by

$$\begin{aligned} f^{\textsf{cr-qpsk}}_{Y}(y) = &\frac{1}{2}\sqrt{\frac{\rho}{\pi}}\left[\frac{Q_2(y_{\textsf{re}})+Q_2(-y_{\textsf{re}})}{\sqrt{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}}\mathrm{e}^{\,-\frac{y_{\textsf{re}}^2\rho}{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}} + Q_1\left(\mathrm{e}^{\,-(y_{\textsf{re}}-1)^2\rho}+\mathrm{e}^{\,-(y_{\textsf{re}}+1)^2\rho}\right)\right] \\ &\cdot\frac{1}{2}\sqrt{\frac{\rho}{\pi}}\left[\frac{Q_2(y_{\textsf{im}})+Q_2(-y_{\textsf{im}})}{\sqrt{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}}\mathrm{e}^{\,-\frac{y_{\textsf{im}}^2\rho}{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}} + Q_1\left(\mathrm{e}^{\,-(y_{\textsf{im}}-1)^2\rho}+\mathrm{e}^{\,-(y_{\textsf{im}}+1)^2\rho}\right)\right] . \end{aligned} \tag{8-6}$$

Finally, following (5-8) and accounting for the inherent symmetry in (8-4) and (8-6), the spectral efficiency of the CR-QPSK scheme is given by

$$\begin{aligned} C^{\textsf{cr-qpsk}}(\rho) = 2\alpha\Bigg(1 - &\int_{-\infty}^{\infty}\sqrt{\frac{\rho}{\pi}}\left[\frac{1}{\sqrt{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}}Q_2(s)\mathrm{e}^{\,-\frac{s^2\rho}{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}} + Q_1\mathrm{e}^{\,-(s-1)^2\rho}\right] \\ &\cdot\log_2\left(1+\frac{Q_2(-s)\mathrm{e}^{\,-\frac{s^2\rho}{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}} + \sqrt{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}Q_1\mathrm{e}^{\,-(s+1)^2\rho}}{Q_2(s)\mathrm{e}^{\,-\frac{s^2\rho}{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}} + \sqrt{1+\rho\alpha\bar{\mathcal{E}}^{\textsf{cr-qpsk}}}Q_1\mathrm{e}^{\,-(s-1)^2\rho}}\right)\mathrm{d}s\Bigg) \quad . \end{aligned} \tag{8-7}$$

Comparative numerical spectral efficiency results are plotted in Figure 11. The figure shows

[10] In the following notation the ± signs are designated with adherence to the corresponding signs of the real and imaginary parts of u. For example, for $u = 1 + j$ one should substitute $Q_2(y_{\textsf{re}})$, $\mathrm{e}^{\,-(y_{\textsf{re}}-1)^2\rho}$, $Q_2(y_{\textsf{im}})$, and $\mathrm{e}^{\,-(y_{\textsf{im}}-1)^2\rho}$ in the corresponding terms in (8-4).

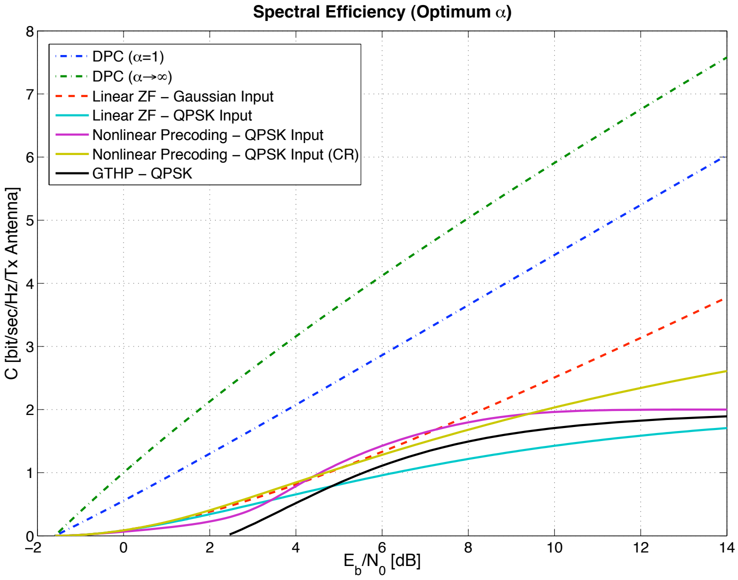

Figure 11: Spectral efficiency results optimized with respect to the load α.

the spectral efficiencies of the discrete extended alphabet relaxation scheme (while taking $L = 2$), and of the CR-QPSK scheme, as well as the spectral efficiency of linear ZF precoding for Gaussian and QPSK input (see (5-15) and (5-17), respectively), and the spectral efficiency of GTHP with QPSK input (following (F-32)–(F-33)). The spectral efficiencies were evaluated for the *optimum* choice of the system load α. The optimum load is a function of $\frac{E_b}{N_0}$ and shown in Figure 12. In Figure 11, the DPC spectral efficiency (5-12) is also provided for comparison, evaluated both for $\alpha = 1$, and for $\alpha \to \infty$ (specifying the ultimate performance). The optimization with respect to α emphasizes its role as a crucial system design parameter, facilitating the proper working point for each transmission scheme, per each $\frac{E_b}{N_0}$. It also naturally translates to a practical scheduling scheme, specifying the desired number of *simultaneously* active scheduled users per transmit antenna (see, e.g., [45]).

The results indicate that nonlinear precoding can provide significant performance enhancement for medium to high $\frac{E_b}{N_0}$ values. The discrete lattice-based relaxation scheme is shown to outperform linear ZF with QPSK input for $\frac{E_b}{N_0} > 3.43\,\text{dB}$. The beneficial effect of the lattice relaxation scheme becomes more pronounced, the more the spectral efficiency approaches the upper limit of 2 bits/sec/Hz per transmit antenna. For example, a spectral efficiency of 1.75 bits/sec/Hz can be obtained with lattice relaxation already at $\frac{E_b}{N_0} \approx 7.66\,\text{dB}$, whereas linear ZF requires additional 7.26 dB for the same spectral efficiency. In fact, the QPSK-based lattice precoding scheme is shown to marginally outperform linear ZF with *Gaussian* input for $4.19\,\text{dB} < \frac{E_b}{N_0} < 7.26\,\text{dB}$. The lattice relaxation scheme also outperforms GTHP for all $\frac{E_b}{N_0}$ values, becoming more effective for medium to high $\frac{E_b}{N_0}$ (for example, GTHP needs 2.93 dB more energy per bit to achieve 1.75

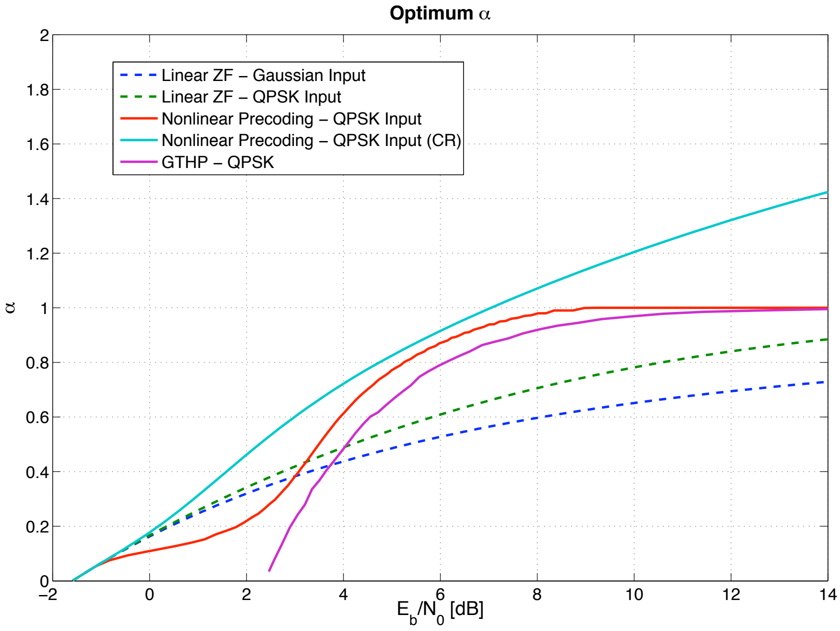

Figure 12: System load that maximizes spectral efficiency as a function of $\frac{E_b}{N_0}$.

bits/sec/Hz)[11]. The gap from the DPC upper bound is, however, still essentially retained (4.49 dB at 1.75 bits/sec/Hz, considering DPC with $\alpha = 1$, to make a fairer comparison).

As for the CR-QPSK scheme, Figure 11 shows that it also provides a considerable performance enhancement over linear ZF with QPSK input. It is outperformed by the lattice relaxation scheme for 4.38 dB $< \frac{E_b}{N_0} <$ 9.40 dB. It performs better at low values of $\frac{E_b}{N_0}$, and in fact it even negligibly outperforms linear ZF with *Gaussian* input in the low $\frac{E_b}{N_0}$ region. Moreover, unlike the discrete scheme, CR-QPSK outperforms linear ZF precoding (with QPSK input) for *all* $\frac{E_b}{N_0}$ values. Furthermore, it outperforms lattice relaxation in the high $\frac{E_b}{N_0}$ region, since it allows for loads up to $\alpha < 2$ and therefore its spectral efficiency is no longer upper bounded by 2 bits/sec/Hz, but rather by 4 bits/sec/Hz per transmit antenna. Though, the convergence to the limiting spectral efficiency of 4 bits/s/Hz at high $\frac{E_b}{N_0}$ is rather slow. CR-QPSK also outperforms GTHP for all $\frac{E_b}{N_0}$ values, but the advantage is more significant for high $\frac{E_b}{N_0}$, where overloading is employed. These results are of particular interest since the CR-QPSK scheme lends itself to efficient implementation, whereas the discrete relaxation scheme involves the solution of an NP-hard optimization problem. It is also important to note that, as shown in Figure 8, the CR-QPSK scheme is always inferior to the lattice relaxation scheme in terms of the limiting energy penalty. Hence, in view of the observations made here, one can conclude that restricting the analysis to the energy penalty alone provides only limited insight into the behavior of large *coded* systems, as it essentially focuses only on the

[11] Note that in general the modulo-receiver employed by GTHP induces poor performance in the low spectral efficiency region (see Appendix F).

transmitter, while ignoring the impact of the nonlinear precoding scheme on the receiver.

9 Concluding Remarks

The replica symmetry breaking ansatz of statistical physics was employed in this paper to investigate the large system limit behavior of nonlinear precoding for the MIMO Gaussian broadcast channel based on linear zero-forcing and alphabet relaxation. For lattice relaxations, the replica symmetric ansatz was shown to yield misleading results for system loads greater than approximately 0.3 while the one-step replica symmetry breaking ansatz provides sensible results for any load. For exact results, however, multiple-step replica symmetry breaking must be considered.

Introducing a nonlinear superchannel comprising the actual channel and the precoder, allows for a Markov chain description of an individual user's channel. This enables the calculation of mutual information and spectral efficiency in the large system limit. While convex QPSK relaxations are significantly outperformed by lattice relaxations in terms of transmitted energy per bit, they are very competitive when combined with strong error-correction coding as shown by the spectral efficiency analysis. Except for medium signal-to-noise ratios, they are superior to lattice relaxations. Both schemes were shown to outperform Tomlinson-Harashima precoding with QPSK input for all signal-to-noise ratios.

The combination of polynomial complexity and high spectral efficiency makes convex alphabet relaxation schemes, as introduced in [24], a promising alternative to the NP-hard lattice relaxations due to their polynomial complexity, and to Tomlinson-Harashima precoding due to their superior performance. The results motivate the search for convex schemes amenable to efficient implementation. Additional examples for extended alphabets are currently investigated, see [25] for preliminary results. Note, however, that the problem of finding the optimum precoding scheme that maximizes the spectral efficiency in this framework is not at all trivial, as the corresponding equivalent channel statistics depend, in this setting, on the choice of input distribution and extended alphabet sets.

A Higher RSB Orders

The r-step RSB ansatz reads

$$\boldsymbol{Q} = q_r \mathbf{1}_{n\times n} + \sum_{i=1}^{r} p_r^{(i)} \boldsymbol{I}_{\frac{n\beta}{\mu_r^{(i)}} \times \frac{n\beta}{\mu_r^{(i)}}} \otimes \mathbf{1}_{\frac{\mu_r^{(i)}}{\beta} \times \frac{\mu_r^{(i)}}{\beta}} + \frac{\chi_r}{\beta} \boldsymbol{I}_{n\times n} \quad , \tag{A-1}$$

using the constants $\left\{q_r, p_r^{(1)}, \ldots, p_r^{(r)}, \chi_r, \mu_r^{(1)}, \ldots, \mu_r^{(r)}\right\}$. The limit as $r \to \infty$ is called *full* RSB and gives the *exact* solution to the problem [37]. The particular temperature-dependent scaling of some parts of $\boldsymbol{Q}$ is used to evaluate the free energy at zero temperature without getting divergent terms. If a finite temperature is of interest a different scaling may be considered. Plugging (A-1) into (3-25), while exploiting the particular structure of $\boldsymbol{Q}$, we find:

Proposition A.1 *For any temperature, the energy for r-step RSB is*

$$\begin{aligned}\mathscr{E}(\beta) = &- q_r \left[\chi_r + \sum_{i=1}^{r} \mu_r^{(i)} p_r^{(i)}\right] R'\left(-\chi_r - \sum_{i=1}^{r} \mu_r^{(i)} p_r^{(i)}\right) + \\ &+ \left[q_r + \frac{\chi_r + \sum_{i=1}^{r} \mu_r^{(i)} p_r^{(i)}}{\mu_r^{(1)}}\right] R\left(-\chi_r - \sum_{i=1}^{r} \mu_r^{(i)} p_r^{(i)}\right) + \\ &+ \sum_{j=2}^{r} \left(\frac{1}{\mu_r^{(j)}} - \frac{1}{\mu_r^{(j-1)}}\right) \left[\chi_r + \sum_{i=j}^{r} \mu_r^{(i)} p_r^{(i)}\right] R\left(-\chi_r - \sum_{i=j}^{r} \mu_r^{(i)} p_r^{(i)}\right) + \\ &+ \left(\frac{\chi_r}{\beta} - \frac{\chi_r}{\mu_r^{(r)}}\right) R(-\chi_r) \quad ,\end{aligned} \tag{A-2}$$

where $R'(\cdot)$ denotes the derivative of the function $R(\cdot)$.

In order to proceed to full RSB, the limit $r \to \infty$ must be taken. Naively, one might think this would make the sums in (A-2) diverge. However, the macroscopic parameters are determined by the saddle point equations, which guarantee that the sums stay finite through decreasing the macroscopic parameters. Thus, we introduce a continuum of macroscopic parameters $\mu(x)$ and $p(x)$, taken over $0 \le x \le 1$, such that

$$q = q_r \ , \tag{A-3}$$

$$\chi = \chi_r \ , \tag{A-4}$$

$$p(i/r) = p_r^{(i)} \ , \tag{A-5}$$

$$\mu(i/r) = \mu_r^{(i)} \ , \tag{A-6}$$

$$p(0) = q_r \ , \tag{A-7}$$

$$\mu(0) = 1 \ , \tag{A-8}$$

and the function

$$\mathscr{G}(x) \triangleq -\chi - \int_x^1 \mu(y) p(y) \mathrm{d}y \quad . \tag{A-9}$$

Accordingly, we find for the energy in the limit $r \to \infty$

$$\begin{aligned}\mathscr{E}(\beta) = q\mathscr{G}(0)\, R'[\mathscr{G}(0)] + \left[q - \frac{\mathscr{G}(0)}{\mu(0)}\right] R[\mathscr{G}(0)] \\ + \int_0^1 \mathscr{G}(x)\, R[\mathscr{G}(x)] \frac{\mathrm{d}\mu(x)}{\mu^2(x)} + \left(\frac{\mathscr{G}(1)}{\mu(1)} + \frac{\chi}{\beta}\right) R[\mathscr{G}(1)] \quad .\end{aligned} \tag{A-10}$$

Using integration by parts, (A-10) simplifies to

$$\mathscr{E}(\beta) = \left. [\mathscr{G}(x)\, R[\mathscr{G}(x)]]' \right|_{x=0} + \frac{\chi}{\beta} R[\mathscr{G}(1)] + \int_0^1 \frac{\mathrm{d}[\mathscr{G}(x)\, R[\mathscr{G}(x)]]}{\mu(x)} \quad . \tag{A-11}$$

The functions $p(x)$ and $\mu(x)$ must be determined by the respective saddle point equations.

B Proofs for 1RSB

B.1 Proposition 4.1

The joint distribution of the entries of the vector $\boldsymbol{x}$, conditioned on both the input vector $\boldsymbol{u}$ and the channel transfer matrix $\boldsymbol{H}$, is given for a non-zero temperature by the Boltzmann distribution

$$P_{\mathcal{B}}(\boldsymbol{x}|\boldsymbol{H},\boldsymbol{u}) = \frac{1}{\mathcal{Z}} \mathrm{e}^{-\beta \boldsymbol{x}^\dagger \boldsymbol{J} \boldsymbol{x}} \quad , \tag{B-1}$$

where $\mathcal{Z}$ is the partition function defined in (3-4). Taking the limit $\beta \to \infty$ (zero temperature), the denominator in (B-1) is dominated by its maximum value term, and the limiting *joint* distribution of the entries of $\boldsymbol{x}$, *conditioned on all inputs*, converges to the Dirac measure at $\operatorname{argmin}_{\boldsymbol{x}\in\mathscr{B}_{\boldsymbol{u}}} \boldsymbol{x}^\dagger \boldsymbol{J} \boldsymbol{x}$, corresponding to the minimum normalized energy penalty, as given by Proposition 4.1.

To prove Proposition 4.1, we will need to evaluate the free energy averaged over all realizations of $\boldsymbol{u}$ and $\boldsymbol{H}$. For future convenience, we also include the dummy variable h and the function $V(\cdot)$ defined in (3-14) and rewrite the free energy as

$$\begin{aligned} -\beta\mathscr{F}(\beta) &= \lim_{K\to\infty} \frac{1}{K} E_{\boldsymbol{u},\boldsymbol{H}} \left\{ \log \mathcal{Z}(h;\boldsymbol{u},\boldsymbol{H}) \right\} \\ &= \lim_{K\to\infty} \sum_{\boldsymbol{u}\in\mathscr{U}^K} P_{U^K}(\boldsymbol{u}) \left(\frac{1}{K} E_{\boldsymbol{H}} \left\{ \log \mathcal{Z}(h;\boldsymbol{u},\boldsymbol{H}) \right\} \right) \quad , \end{aligned} \tag{B-2}$$

where $\mathcal{Z}(h;\boldsymbol{u},\boldsymbol{H})$ is given by (3-15). The second equality is a manifestation of the underlying assumption that the coded symbols of all users are drawn randomly and independently of the channel transfer matrix $\boldsymbol{H}$. In view of this formulation, we consider now the limit of the term in the parentheses above

$$\lim_{K\to\infty} \frac{1}{K} E_{\boldsymbol{H}} \left\{ \log \mathcal{Z}(h;\boldsymbol{u},\boldsymbol{H}) \right\} \quad . \tag{B-3}$$

As shown later on, this inner limit is a deterministic quantity, for almost every realization of the input vector $\boldsymbol{u}$. It will hence be concluded that in fact

$$-\beta\mathscr{F}(\beta) = \lim_{K\to\infty} \frac{1}{K} E_{\boldsymbol{H}} \left\{ \log \mathcal{Z}(h;\boldsymbol{u},\boldsymbol{H}) \right\} \quad . \tag{B-4}$$

As indicated earlier, the key tool in the derivation of the above quantity is the replica method of statistical physics, using the identity

$$E_{\boldsymbol{H}} \left\{ \log \mathcal{Z}(h;\boldsymbol{u},\boldsymbol{H}) \right\} = \lim_{n\to 0} \frac{1}{n} \log E_{\boldsymbol{H}} \left\{ [\mathcal{Z}(h;\boldsymbol{u},\boldsymbol{H})]^n \right\} \quad , \tag{B-5}$$

and following the outline in Section 3. With that in mind, the quantity $[\mathcal{Z}(h;\boldsymbol{u},\boldsymbol{H})]^n$ is regarded as consisting of n identical replicas of the original (unnormalized) probability model in the following way [28]

$$
\begin{aligned}
[\mathcal{Z}(h;\boldsymbol{u},\boldsymbol{H})]^{n} &= \left(\sum_{\boldsymbol{x}\in\mathscr{B}_{\boldsymbol{u}}} \mathrm{e}^{-\beta V(h,\xi,\upsilon,\boldsymbol{x},\boldsymbol{u}))}\mathrm{e}^{-\beta\boldsymbol{x}^{\dagger}\boldsymbol{J}\boldsymbol{x}}\right)^{n} \\
&= \sum_{\{\boldsymbol{x}_a\}} \mathrm{e}^{-\beta\sum_{a=1}^{n}V(h,\xi,\upsilon,\boldsymbol{x}_a,\boldsymbol{u})}\mathrm{e}^{\sum_{a=1}^{n}-\beta\boldsymbol{x}_a^{\dagger}\boldsymbol{J}\boldsymbol{x}_a} \quad .
\end{aligned}
\tag{B-6}
$$

Interchanging the limits of $K\to\infty$ and $n\to 0$, the focus is first on the derivation of the limit

$$
\begin{aligned}
\Xi_n &\triangleq \lim_{K\to\infty}\frac{1}{K}\log E_{\boldsymbol{H}}\left\{[\mathcal{Z}(h;\boldsymbol{u},\boldsymbol{H})]^{n}\right\} \\
&= \lim_{K\to\infty}\frac{1}{K}\log E_{\boldsymbol{H}}\left\{\sum_{\{\boldsymbol{x}_a\}} \mathrm{e}^{-\beta\sum_{a=1}^{n}V(h,\xi,\upsilon,\boldsymbol{x}_a,\boldsymbol{u})}\mathrm{e}^{-\operatorname{Tr}(\beta\boldsymbol{J}\sum_{a=1}^{n}\boldsymbol{x}_a\boldsymbol{x}_a^{\dagger})}\right\} \quad .
\end{aligned}
\tag{B-7}
$$

Since the first exponential term within the expectation is independent of the channel transfer matrix $\boldsymbol{H}$, Ξ_n can be rewritten as

$$
\Xi_n = \lim_{K\to\infty}\frac{1}{K}\log\left(\sum_{\{\boldsymbol{x}_a\}} \mathrm{e}^{-\beta\sum_{a=1}^{n}V(h,\xi,\upsilon,\boldsymbol{x}_a,\boldsymbol{u})}E_{\boldsymbol{H}}\left\{\mathrm{e}^{-\operatorname{Tr}(\beta\boldsymbol{J}\sum_{a=1}^{n}\boldsymbol{x}_a\boldsymbol{x}_a^{\dagger})}\right\}\right) \quad . \tag{B-8}
$$

The inner expectation in (B-8) is the Harish-Chandra-Itzykson-Zuber integral (see [24] and references therein), and the objective here is its evaluation for fixed-rank matrices $\sum_{a=1}^{n}\boldsymbol{x}_a\boldsymbol{x}_a^{\dagger}$, in the large K limit. This problem was recently considered in [53], and invoking Theorem 1.7 therein, (B-8) can be represented for large K as[12]

$$
\Xi_n = \lim_{K\to\infty}\frac{1}{K}\log\left(\sum_{\{\boldsymbol{x}_a\}} \mathrm{e}^{-\beta\sum_{a=1}^{n}V(h,\xi,\upsilon,\boldsymbol{x}_a,\boldsymbol{u})}\mathrm{e}^{-K\sum_{a=1}^{n}\int_0^{\lambda_a}R(-w)\,\mathrm{d}w+o(K)}\right) \quad , \tag{B-9}
$$

where $R(w)$ is the R-transform of the limiting eigenvalue distribution of the matrix $\boldsymbol{J}$, and $\{\lambda_a\}$ denote the eigenvalues of the $n\times n$ matrix $\beta\boldsymbol{Q}$ with $\boldsymbol{Q}$ defined through[13]

$$
Q_{ab} = \frac{1}{K}\boldsymbol{x}_a^{\dagger}\boldsymbol{x}_b \triangleq \frac{1}{K}\sum_{k=1}^{K}x_{ak}^{*}x_{bk} \quad . \tag{B-10}
$$

Since additive exponential terms of order $o(K)$ have no effect on the results in the limiting regime as $K\to\infty$, due to the $\frac{1}{K}$ factor outside the logarithm in (B-9) (this shall become clear in the derivation to follow), any such terms are dropped henceforth for notational simplicity.

In order to calculate the summation in (B-9), the procedure employed in [24] is repeated here, and the Kn-dimensional space spanned by the replicas is split into subshells by means of (3-24). Assuming $\boldsymbol{Q}^{\dagger}=\boldsymbol{Q}$, Ξ_n can be represented as

$$
\Xi_n = \lim_{K\to\infty}\frac{1}{K}\log\left(\int \mathrm{e}^{K\text{Ł}}\mathrm{e}^{K\mathcal{I}(\boldsymbol{Q})}\mathrm{e}^{-K\mathcal{G}(\boldsymbol{Q})}\,\mathcal{D}\boldsymbol{Q}\right) \quad , \tag{B-11}
$$

[12] $o(K)$ is used here to denote quantities that satisfy $\lim_{K\to\infty}o(K)/K=0$.

[13] Here [53, Theorem 1.7] is applied individually for all *given* vectors $\{\boldsymbol{x}_a\}$.

where

$$\mathcal{D}\boldsymbol{Q} = \prod_{a=1}^{n} \mathrm{d}Q_{aa} \prod_{b=a+1}^{n} \mathrm{d}\Re(Q_{ab})\, \mathrm{d}\Im(Q_{ab}) \tag{B-12}$$

is the integration measure,

$$\begin{aligned} \mathcal{G}(\boldsymbol{Q}) &= \sum_{a=1}^{n} \int\limits_{0}^{\beta\lambda_a(\boldsymbol{Q})} R(-w)\,\mathrm{d}w && \text{(B-13)} \\ &= \sum_{a=1}^{n} \int\limits_{0}^{\beta} \lambda_a(\boldsymbol{Q}) R(-w\lambda_a(\boldsymbol{Q}))\,\mathrm{d}w && \text{(B-14)} \\ &= \int\limits_{0}^{\beta} \mathrm{Tr}\left[\boldsymbol{Q}R(-w\boldsymbol{Q})\right]\,\mathrm{d}w \quad , && \text{(B-15)} \end{aligned}$$

since the trace is the sum of the eigenvalues,

$$\text{\L} = -\frac{\beta}{K} \sum_{a=1}^{n} V(h, \xi, \upsilon, \boldsymbol{x}_a, \boldsymbol{u}) \quad , \tag{B-16}$$

and

$$\mathrm{e}^{\,K\mathcal{I}(\boldsymbol{Q})} = \sum_{\{\boldsymbol{x}_a\}} \prod_{a=1}^{n} \delta(\boldsymbol{x}_a^{\dagger}\boldsymbol{x}_a - KQ_{aa}) \prod_{b=a+1}^{n} \delta(\Re[\boldsymbol{x}_a^{\dagger}\boldsymbol{x}_b - KQ_{ab}])\, \delta(\Im[\boldsymbol{x}_a^{\dagger}\boldsymbol{x}_b - KQ_{ab}]) \tag{B-17}$$

is the probability weight of the subshell.

Starting with $\mathrm{e}^{\,K\mathcal{I}(\boldsymbol{Q})}\mathrm{e}^{\,K\text{\L}}$ we follow [24] and represent the Dirac measures using the inverse Laplace transform. This is performed by introducing the *complex* variables $\left\{\tilde{Q}_{ab}^{(I)}\right\}$, $1 \le a \le b \le n$, and $\left\{\tilde{Q}_{ab}^{(Q)}\right\}$, $1 \le a \le b \le n$, and defining the matrix $\tilde{\boldsymbol{Q}}$ with elements (taking $a < b$)

$$\begin{aligned} \tilde{Q}_{aa} &= \tilde{Q}_{aa}^{(I)} \quad , && \text{(B-18)} \\ \tilde{Q}_{ab} &= \frac{1}{2}\left(\tilde{Q}_{ab}^{(I)} - j\tilde{Q}_{ab}^{(Q)}\right) \quad , && \text{(B-19)} \\ \tilde{Q}_{ba} &= \frac{1}{2}\left(\tilde{Q}_{ab}^{(I)} + j\tilde{Q}_{ab}^{(Q)}\right) \quad . && \text{(B-20)} \end{aligned}$$

Denoting by $\boldsymbol{P}$ the Hermitian matrix with elements $P_{ab} = \boldsymbol{x}_a^{\dagger}\boldsymbol{x}_b - KQ_{ab}$, this yields

$$\begin{aligned} \delta(P_{aa}) &= \int_{\mathcal{J}} \mathrm{e}^{\,\tilde{Q}_{aa}P_{aa}} \frac{d\tilde{Q}_{aa}^{(I)}}{2\pi j} \quad , && \text{(B-21)} \\ \delta(\Re\{P_{ab}\})\,\delta(\Im\{P_{ab}\}) &= \int_{\mathcal{J}^2} \mathrm{e}^{\,\tilde{Q}_{ab}^{(I)}\Re\{P_{ab}\} - \tilde{Q}_{ab}^{(Q)}\Im\{P_{ab}\}} \frac{d\tilde{Q}_{ab}^{(I)}\, d\tilde{Q}_{ab}^{(Q)}}{(2\pi j)^2} && \text{(B-22)} \\ &= \int_{\mathcal{J}^2} \mathrm{e}^{\,\tilde{Q}_{ab}P_{ba} + \tilde{Q}_{ba}P_{ab}} \frac{d\tilde{Q}_{ab}^{(I)}\, d\tilde{Q}_{ab}^{(Q)}}{(2\pi j)^2} \quad , && \text{(B-23)} \end{aligned}$$

where the integration is over $\mathcal{J} = (t - j\infty, t + j\infty)$, for some $t \in \mathbb{R}$ (note that $P_{ab} = P_{ba}^{*}$).

Substituting in (B-17) and combining with (B-16), it follows that

$$\begin{aligned}
\mathrm{e}^{K\mathcal{I}(\boldsymbol{Q})}\mathrm{e}^{K\text{Ł}} &= \sum_{\{\boldsymbol{x}_a\}} \int_{\mathcal{J}^{n^2}} \mathrm{e}^{\sum_{a,b} \tilde{Q}_{ab}(\boldsymbol{x}_b^\dagger \boldsymbol{x}_a - KQ_{ba})} \mathrm{e}^{-\beta \sum_{a=1}^{n} V(h,\xi,\upsilon,\boldsymbol{x}_a,\boldsymbol{u})} \tilde{\mathfrak{D}}\tilde{\boldsymbol{Q}} \\
&= \int_{\mathcal{J}^{n^2}} \mathrm{e}^{-K\,\mathrm{Tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q})} \left(\sum_{\{\boldsymbol{x}_a\}} \mathrm{e}^{\sum_{a,b} \tilde{Q}_{ab}\boldsymbol{x}_b^\dagger \boldsymbol{x}_a} \mathrm{e}^{-\beta \sum_{a=1}^{n} V(h,\xi,\upsilon,\boldsymbol{x}_a,\boldsymbol{u})} \right) \tilde{\mathfrak{D}}\tilde{\boldsymbol{Q}} \quad ,
\end{aligned} \tag{B-24}$$

where

$$\tilde{\mathfrak{D}}\tilde{\boldsymbol{Q}} = \prod_{a=1}^{n} \left(\frac{d\tilde{Q}_{aa}^{(I)}}{2\pi j} \prod_{b=a+1}^{n} \frac{d\tilde{Q}_{ab}^{(I)}\, d\tilde{Q}_{ab}^{(Q)}}{(2\pi j)^2} \right) . \tag{B-25}$$

Considering the inner summation in (B-24), then rearranging terms and using (3-14) the expression can be rewritten as

$$\sum_{\{\boldsymbol{x}_a\}} \mathrm{e}^{\sum_{a,b} \tilde{Q}_{ab}\boldsymbol{x}_b^\dagger \boldsymbol{x}_a} \mathrm{e}^{-\beta \sum_{a=1}^{n} V(h,\xi,\upsilon,\boldsymbol{x}_a,\boldsymbol{u})} = \prod_{k=1}^{K} \sum_{\left\{x_a \in \mathscr{B}_{u_k}\right\}} \mathrm{e}^{\left(\sum_{a,b} \tilde{Q}_{ab} x_b^* x_a\right) + h\beta \sum_{a=1}^{n} 1\{(x_a,u_k)=(\xi,\upsilon)\}} \quad . \tag{B-26}$$

Defining

$$M_k(\tilde{\boldsymbol{Q}}) = \sum_{\left\{x_a \in \mathscr{B}_{u_k}\right\}} \mathrm{e}^{\left(\sum_{a,b} x_b^* x_a \tilde{Q}_{ab}\right) + h\beta \sum_{a=1}^{n} 1\{(x_a,u_k)=(\xi,\upsilon)\}} \quad , \tag{B-27}$$

one finally gets

$$\mathrm{e}^{K\mathcal{I}(\boldsymbol{Q})}\mathrm{e}^{K\text{Ł}} = \int_{\mathcal{J}^{n^2}} \mathrm{e}^{-K\,\mathrm{Tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q}) + \sum_{k=1}^{K} \log M_k(\tilde{\boldsymbol{Q}})} \tilde{\mathfrak{D}}\tilde{\boldsymbol{Q}} \quad . \tag{B-28}$$

Now, using the underlying assumption that the coded symbols transmitted by different users are i.i.d., one can apply the strong law of large numbers for $K \to \infty$ to get

$$\log M(\tilde{\boldsymbol{Q}}) \triangleq \frac{1}{K} \sum_{k=1}^{K} \log M_k(\tilde{\boldsymbol{Q}}) \tag{B-29}$$

$$\to \int \log \sum_{\{x_a \in \mathscr{B}_u\}} \mathrm{e}^{\left(\sum_{a,b} x_b^* x_a \tilde{Q}_{ab}\right) + h\beta \sum_{a=1}^{n} 1\{(x_a,u)=(\xi,\upsilon)\}} \mathrm{d}F_U(u) \tag{B-30}$$

$$= \int \log \sum_{\mathbf{x} \in \mathscr{B}_u^n} \mathrm{e}^{\mathbf{x}^\dagger \tilde{\boldsymbol{Q}} \mathbf{x} + h\beta \sum_{a=1}^{n} 1\{(x_a,u)=(\xi,\upsilon)\}} \mathrm{d}F_U(u) \ , \tag{B-31}$$

where the convergence is in the *almost sure* sense, for any extended alphabets such that the expectation in (B-30) exists. Note that this observation implies that any randomness due to $\boldsymbol{u}$ in the RHS of (B-11) effectively vanishes at the large system limit, due to the normalization with respect to K outside the logarithm.

The next step in the evaluation of (B-11) is the observation that in the limit as $K \to \infty$, the integrand therein is dominated by the exponential term with maximal exponent. Therefore, only the subshell that corresponds to this extremal value of the correlation between the vectors $\{\boldsymbol{x}_a\}$

is relevant for the calculation of the integral. Thus, we have at the saddle point

$$\frac{\partial}{\partial \boldsymbol{Q}} \left[\mathcal{G}(\boldsymbol{Q}) + \mathrm{Tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q}) \right] = \mathbf{0} \quad . \tag{B-32}$$

Since the trace is the sum of the eigenvalues, we can write (B-13) as

$$\mathcal{G}(\boldsymbol{Q}) = \mathrm{Tr} \int\limits_0^{\beta \boldsymbol{Q}} R(-w)\,\mathrm{d}w \tag{B-33}$$

and (B-32) gives

$$\tilde{\boldsymbol{Q}} = -\beta R(-\beta \boldsymbol{Q}) \quad . \tag{B-34}$$

Furthermore, we observe that also the integrand in (B-28) is dominated by the exponential term with maximal exponent in the limit $K \to \infty$. Thus, at the saddle point we have

$$\frac{\partial}{\partial \tilde{\boldsymbol{Q}}} \left[\log M(\tilde{\boldsymbol{Q}}) - \mathrm{Tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q}) \right] = \mathbf{0} \quad . \tag{B-35}$$

With (B-31), this gives

$$\boldsymbol{Q} = \int \frac{\sum\limits_{\mathbf{x} \in \mathscr{B}_u^n} \mathbf{x}\mathbf{x}^\dagger \mathrm{e}^{\mathbf{x}^\dagger \tilde{\boldsymbol{Q}} \mathbf{x} + h\beta \sum\limits_{a=1}^{n} 1\{(x_a, u) = (\xi, \upsilon)\}}}{\sum\limits_{\mathbf{x} \in \mathscr{B}_u^n} \mathrm{e}^{\mathbf{x}^\dagger \tilde{\boldsymbol{Q}} \mathbf{x} + h\beta \sum\limits_{a=1}^{n} 1\{(x_a, u) = (\xi, \upsilon)\}}} \,\mathrm{d}F_U(u) \quad . \tag{B-36}$$

We now invoke the *1RSB* assumption (3-31) regarding the structure of the matrices $\boldsymbol{Q}$ at the saddle-point that dominate the integral. In a similar manner we set

$$\tilde{\boldsymbol{Q}} = \beta^2 f_1^2 \mathbf{1}_{n \times n} + \beta^2 g_1^2 \boldsymbol{I}_{\frac{n\beta}{\mu_1} \times \frac{n\beta}{\mu_1}} \otimes \mathbf{1}_{\frac{\mu_1}{\beta} \times \frac{\mu_1}{\beta}} - \beta \varepsilon_1 \boldsymbol{I}_{n \times n} \quad , \tag{B-37}$$

introducing the macroscopic parameters f_1, g_1, and ε_1.

With these assumptions one can explicitly obtain the eigenvalues of the matrix $\beta \boldsymbol{Q}$ [14], and $\mathcal{G}(\boldsymbol{Q})$ can be rewritten as

$$\begin{aligned} \mathcal{G}(q_1, p_1, \chi_1, \mu_1) = \left(n - \frac{n\beta}{\mu_1}\right) \int\limits_0^{\chi_1} R(-w)\,\mathrm{d}w \\ + \left(\frac{n\beta}{\mu_1} - 1\right) \int\limits_0^{\chi_1 + \mu_1 p_1} R(-w)\,\mathrm{d}w + \int\limits_0^{\chi_1 + \mu_1 p_1 + \beta n q_1} R(-w)\,\mathrm{d}w \ . \end{aligned} \tag{B-38}$$

[14] The eigenvalue $(\beta n q_1 + \mu_1 p + \chi_1)$ occurs with multiplicity 1, the eigenvalue $(\mu_1 p_1 + \chi_1)$ occurs with multiplicity $(\frac{n\beta}{\mu_1} - 1)$, and the eigenvalue χ_1 occurs with multiplicity $(n - \frac{n\beta}{\mu_1})$.

It also follows from the 1RSB assumption that

$$\mathrm{Tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q}) = \begin{bmatrix}\beta^2 f_1^2 & \beta^2 g_1^2 & -\beta\varepsilon_1\end{bmatrix} \begin{bmatrix} n^2 & \frac{n\mu_1}{\beta} & n \\ \frac{n\mu_1}{\beta} & \frac{n\mu_1}{\beta} & n \\ n & n & n \end{bmatrix} \begin{bmatrix} q_1 \\ p_1 \\ \frac{\chi_1}{\beta} \end{bmatrix} \quad , \tag{B-39}$$

and

$$\log M(f_1, g_1, \varepsilon_1, \mu_1) =$$
$$\int \log \sum_{\{x_a \in \mathscr{B}_u\}} \mathrm{e}^{\beta^2 f_1^2 \left|\sum\limits_{a=1}^{n} x_a\right|^2 + \beta^2 g_1^2 \sum\limits_{l=0}^{\frac{n\beta}{\mu_1}-1} \left|\sum\limits_{a=1}^{\frac{\mu}{\beta}} x_{a+\frac{l\mu_1}{\beta}}\right|^2 - \beta\varepsilon_1 \sum\limits_{a=1}^{n} |x_a|^2 + h\beta \sum\limits_{a=1}^{n} 1\{(x_a, u) = (\xi, v)\}} \mathrm{d}F_U(u) \ . \tag{B-40}$$

Due to (B-32), the partial derivatives of

$$\mathcal{G}(q_1, p_1, \chi_1, \mu_1) + \mathrm{Tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q}) \tag{B-41}$$

with respect to q_1, p_1, and χ_1 must vanish as $K \to \infty$ by definition of the saddle point. Using (B-38) and (B-39) this yields the following set of equations

$$0 = n^2\beta^2 f_1^2 + n\beta\mu_1 g_1^2 - n\beta\varepsilon_1 + n\beta R(-\chi_1 - \mu_1 p_1 - \beta n q_1) \ , \tag{B-42}$$

$$\begin{aligned} 0 &= n\beta\mu_1 f_1^2 + n\beta\mu_1 g_1^2 - n\beta\varepsilon_1 + (n\beta - \mu_1) R(-\chi_1 - \mu_1 p_1) \\ &\quad + \mu_1 R(-\chi_1 - \mu_1 p_1 - n\beta q_1) \ , \end{aligned} \tag{B-43}$$

$$\begin{aligned} 0 &= n\beta f_1^2 + n\beta g_1^2 - n\varepsilon_1 + \left(n - \frac{n\beta}{\mu_1}\right) R(-\chi_1) + \left(\frac{n\beta}{\mu_1} - 1\right) R(-\chi_1 - \mu_1 p_1) \\ &\quad + R(-\chi_1 - \mu_1 p_1 - n\beta q_1) \ . \end{aligned} \tag{B-44}$$

Solving for ε_1, g_1, and f_1, while focusing on the limit as $n \to 0$, one gets

$$\varepsilon_1 = R(-\chi_1) \ , \tag{B-45}$$

$$g_1 = \sqrt{\frac{R(-\chi_1) - R(-\chi_1 - \mu_1 p_1)}{\mu_1}} \ , \tag{B-46}$$

$$f_1 = \sqrt{\frac{R(-\chi_1 - \mu_1 p_1) - R(-\chi_1 - \mu_1 p_1 - n\beta q_1)}{n\beta}} \xrightarrow{n \to 0} \sqrt{q_1 R'(-\chi_1 - \mu_1 p_1)} \ . \tag{B-47}$$

We now rewrite the expression for $M_k(f_1, g_1, \varepsilon_1, \mu_1)$ in (B-40) using the Hubbard-Stratonovich transform and the shortened notation of (4-6)

$$\mathrm{e}^{|x|^2} = \int\limits_{\mathbb{C}} \mathrm{e}^{2\Re\{xz^*\}} \, \mathrm{D}z \quad , \tag{B-48}$$

yielding (c.f. [24, (66)-(70)])

$$\log M(f_1, g_1, \varepsilon_1, \mu_1) =$$

$$= \int \log \sum_{\{x_a \in \mathscr{B}_u\}} \int_{\mathbb{C}} \mathrm{e}^{\sum_{a=1}^{n}\left[2\beta f_1 \Re\{x_a z^*\} - \beta\varepsilon_1 |x_a|^2 + h\beta\, 1\{(x_a,u)=(\xi,\upsilon)\}\right] + \beta^2 g_1^2 \sum_{l=0}^{\frac{n\beta}{\mu_1}-1} \left|\sum_{a=1}^{\frac{\mu_1}{\beta}} x_{a+\frac{l\mu_1}{\beta}}\right|^2} \mathrm{D}z \, \mathrm{d}F_U(u) \tag{B-49}$$

$$= \int \log \int_{\mathbb{C}} \left[\int_{\mathbb{C}} \left(\sum_{x \in \mathscr{B}_u} \mathcal{K}(u, x, y, z) \right)^{\frac{\mu_1}{\beta}} \mathrm{D}y \right]^{\frac{n\beta}{\mu_1}} \mathrm{D}z \, \mathrm{d}F_U(u) \ , \tag{B-50}$$

with

$$\mathcal{K}(u, x, y, z) \triangleq \mathrm{e}^{\,2\beta \Re\{x(f_1 z^* + g_1 y^*)\} - \beta\varepsilon_1 |x|^2 + h\beta\, 1\{(x,u)=(\xi,\upsilon)\}} \quad . \tag{B-51}$$

Due to (B-35), the partial derivatives of

$$\log M(f_1, g_1, \varepsilon_1, \mu_1) - \mathrm{Tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q}) \tag{B-52}$$

with respect to f_1, g_1 and ε_1, must also vanish as $K \to \infty$. This produces the following set of equations (while taking the limit as $n \to 0$)

$$\chi_1 + p_1\mu_1 = \frac{1}{f_1} \int\!\!\int_{\mathbb{C}^2} \frac{\left(\sum_{x \in \mathscr{B}_u} \mathcal{K}(u, x, y, z) \right)^{\frac{\mu_1}{\beta}-1}}{\int_{\mathbb{C}} \left(\sum_{x \in \mathscr{B}_u} \mathcal{K}(u, x, \tilde{y}, z) \right)^{\frac{\mu_1}{\beta}} \mathrm{D}\tilde{y}} \sum_{x \in \mathscr{B}_u} \Re\{xz^*\} \, \mathcal{K}(u, x, y, z) \, \mathrm{D}y \, \mathrm{D}z \, \mathrm{d}F_U(u) \ , \tag{B-53}$$

$$\chi_1 + (q_1 + p_1)\mu_1 = \frac{1}{g_1} \int\!\!\int_{\mathbb{C}^2} \frac{\left(\sum_{x \in \mathscr{B}_u} \mathcal{K}(u, x, y, z) \right)^{\frac{\mu_1}{\beta}-1}}{\int_{\mathbb{C}} \left(\sum_{x \in \mathscr{B}_u} \mathcal{K}(u, x, \tilde{y}, z) \right)^{\frac{\mu_1}{\beta}} \mathrm{D}\tilde{y}} \cdot \sum_{x \in \mathscr{B}_u} \Re\{xy^*\} \, \mathcal{K}(u, x, y, z) \, \mathrm{D}y \, \mathrm{D}z \, \mathrm{d}F_U(u) \ , \tag{B-54}$$

$$q_1 + p_1 = \iint\limits_{\mathbb{C}^2} \frac{\left(\sum\limits_{x\in\mathscr{B}_u} \mathcal{K}(u,x,y,z)\right)^{\frac{\mu_1}{\beta}-1}}{\int\limits_{\mathbb{C}} \left(\sum\limits_{x\in\mathscr{B}_u} \mathcal{K}(u,x,\tilde{y},z)\right)^{\frac{\mu_1}{\beta}} \mathrm{D}\tilde{y}} \sum_{x\in\mathscr{B}_u} |x|^2\, \mathcal{K}(u,x,y,z)\, \mathrm{D}y\, \mathrm{D}z\, \mathrm{d}F_U(u) - \frac{\chi_1}{\beta} \quad . \tag{B-55}$$

The parameter μ_1 should also be chosen such that the partial derivative of

$$\mathcal{G}(q_1,p_1,\chi_1,\mu_1) + \mathrm{Tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q}) - \log M(f_1,g_1,\varepsilon_1,\mu_1) \tag{B-56}$$

with respect to μ_1 vanishes. This yields at the limit as $n \to 0$

$$\begin{aligned} 0 = -\frac{1}{\mu_1^2}\int\limits_{\chi_1}^{\chi_1+\mu_1 p_1} R(-w)\, dw + \frac{p_1}{\mu_1} R(-\chi_1) + q_1 g_1^2 + \iint\limits_{\mathbb{C}} \Bigg[\frac{1}{\mu_1^2} \log\left(\int\limits_{\mathbb{C}} \left(\sum_{x\in\mathscr{B}_u} \mathcal{K}(u,x,y,z) \right)^{\frac{\mu_1}{\beta}} \mathrm{D}y \right) \\ - \int\limits_{\mathbb{C}} \frac{\left(\sum\limits_{x\in\mathscr{B}_u} \mathcal{K}(u,x,y,z)\right)^{\frac{\mu_1}{\beta}}}{\beta\mu_1 \int\limits_{\mathbb{C}} \left(\sum\limits_{x\in\mathscr{B}_u} \mathcal{K}(u,x,\tilde{y},z)\right)^{\frac{\mu_1}{\beta}} \mathrm{D}\tilde{y}} \cdot \log\left(\sum_{x\in\mathscr{B}_u} \mathcal{K}(u,x,y,z) \right) \mathrm{D}y \Bigg] \mathrm{D}z\, \mathrm{d}F_U(u) \quad . \end{aligned} \tag{B-57}$$

Incorporating all previous results, we get that the quantity Ξ_n of (B-7) is equal to

$$\begin{aligned} \Xi_n &= \mathcal{I}(\boldsymbol{Q}) + \text{Ł} - \mathcal{G}(\boldsymbol{Q}) \\ &= \log M(f_1,g_1,\varepsilon_1,\mu_1) - \beta^2 f_1^2 q_1 n^2 \\ &\quad - \left(\beta f_1^2 (p_1\mu_1 + \chi_1) + \beta g_1^2 (q_1\mu_1 + p_1\mu_1 + \chi_1) - \beta\varepsilon_1 (q_1 + p_1 + \frac{\chi_1}{\beta}) \right) n \\ &\quad - \left(n - \frac{n\beta}{\mu_1} \right) \int\limits_0^{\chi_1} R(-w)\, \mathrm{d}w - \left(\frac{n\beta}{\mu_1} - 1 \right) \int\limits_0^{\chi_1+\mu_1 p_1} R(-w)\, \mathrm{d}w \\ &\quad - \int\limits_0^{\chi_1+\mu_1 p_1+\beta n q_1} R(-w)\, \mathrm{d}w \quad , \end{aligned} \tag{B-58}$$

where the macroscopic parameters $\{f_1, g_1, \varepsilon_1, q_1, p_1, \chi_1, \mu_1\}$ are obtained from the saddle point fixed-point equations (B-45)–(B-47), (B-53)–(B-55), and (B-57). Now in view of (B-5), the next step in the derivation is to take the limit

$$\lim_{n\to 0}\frac{1}{n}\Xi_n = \frac{\beta}{\mu_1}\int\int\limits_{\mathbb{C}} \log\left(\int\limits_{\mathbb{C}}\left(\sum_{x\in\mathscr{B}_u}\mathcal{K}(u,x,y,z)\right)^{\frac{\mu_1}{\beta}} \mathrm{D}y\right)\mathrm{D}z\,\mathrm{d}F_U(u)$$
$$-\beta f_1^2(\chi_1+p_1\mu_1)-\beta g_1^2(\chi_1+(p_1+q_1)\mu_1)+\beta\varepsilon_1(q_1+p_1+\frac{\chi_1}{\beta})$$
$$-\left(1-\frac{\beta}{\mu_1}\right)\int\limits_0^{\chi_1} R(-w)\,\mathrm{d}w-\frac{\beta}{\mu_1}\int\limits_0^{\chi_1+\mu_1 p_1} R(-w)\,\mathrm{d}w-\beta q_1 R(-\chi_1-\mu_1 p_1)\quad . \tag{B-59}$$

But in fact

$$\lim_{n\to 0}\frac{1}{n}\Xi_n = -\beta\mathscr{F}(\beta,h)\quad , \tag{B-60}$$

which is justified by the observation that Ξ_n converges to the same limit for almost every realization of $\boldsymbol{u}$, applying the law of large numbers in (B-30) (see also (B-3) and the discussion that follows).

We note at this point that the energy penalty $\bar{\mathscr{E}}_{\mathsf{rsb1}}$ satisfies

$$\bar{\mathscr{E}}_{\mathsf{rsb1}} = \lim_{\beta\to\infty}\mathscr{F}(\beta,h)|_{h=0}\quad , \tag{B-61}$$

and (4-12) can be readily expressed from Proposition A.1 as a function of the macroscopic parameters $\{q_1,p_1,\chi_1,\mu_1\}$. In order to evaluate the energy penalty, it is thus left to derive the fixed point equations that determine these parameters, as given by (4-8)–(4-11), which is obtained by substituting $h=0$ and taking the limit as $\beta\to\infty$ in (B-53)–(B-55), and (B-57) after back-substitution of (B-51). This completes the proof of Proposition 4.1.

B.2 Proposition 4.2

We derive the limiting conditional distribution of the precoder output x given the input u starting from (3-17). We therefore need to evaluate the derivative of the free energy with respect to h. This can be done directly given (B-60). Taking the partial derivative in (B-59) and using (B-51), we get

$$P_{X|U}(\xi|\upsilon) = \frac{1}{P_U(\upsilon)}\left.\frac{\partial}{\partial h}\mathscr{F}(\beta,h)\right|_{h=0} \tag{B-62}$$

$$= \int\int_{\mathbb{C}^2}\frac{\left(\sum_{x\in\mathscr{B}_u}\mathrm{e}^{2\beta\Re\{x(f_1z^*+g_1y^*)\}-\beta\varepsilon_1|x|^2}\right)^{\frac{\mu_1}{\beta}-1}}{\int_{\mathbb{C}}\left(\sum_{x\in\mathscr{B}_u}\mathrm{e}^{2\beta\Re\{x(f_1z^*+g_1\tilde{y}^*)\}-\beta\varepsilon_1|x|^2}\right)^{\frac{\mu_1}{\beta}}\mathrm{D}\tilde{y}}$$
$$\cdot\sum_{x\in\mathscr{B}_u}\mathrm{e}^{2\beta\Re\{x(f_1z^*+g_1y^*)\}-\beta\varepsilon_1|x|^2}\,\mathrm{D}y\,\mathrm{D}z\,\frac{1\{(x,u)=(\xi,\upsilon)\}}{P_U(\upsilon)}\,\mathrm{d}F_U(u) \tag{B-63}$$

$$= \int_{\mathbb{C}^2}\frac{\left(\sum_{x\in\mathscr{B}_\upsilon}\mathrm{e}^{2\beta\Re\{x(f_1z^*+g_1y^*)\}-\beta\varepsilon_1|x|^2}\right)^{\frac{\mu_1}{\beta}-1}}{\int_{\mathbb{C}}\left(\sum_{x\in\mathscr{B}_\upsilon}\mathrm{e}^{2\beta\Re\{x(f_1z^*+g_1\tilde{y}^*)\}-\beta\varepsilon_1|x|^2}\right)^{\frac{\mu_1}{\beta}}\mathrm{D}\tilde{y}}\,\mathrm{e}^{2\beta\Re\{\xi(f_1z^*+g_1y^*)\}-\beta\varepsilon_1|\xi|^2}\,\mathrm{D}y\,\mathrm{D}z\ . \tag{B-64}$$

After taking the limit $\beta \to \infty$, while applying the saddle point integration rule, we finally get (4-13). This completes the proof of Proposition 4.2.

B.3 Proposition 4.5

We start from (3-10), (3-19), and (3-20) which yield

$$\begin{aligned} \mathscr{S}(\beta) &= \beta \frac{\mathrm{d}\left(\beta \mathscr{F}(\beta)\right)}{\mathrm{d}\beta} - \beta \mathscr{F}(\beta) \\ &= \beta \frac{\partial \left(\beta \mathscr{F}(\beta)\right)}{\partial \beta} - \beta \mathscr{F}(\beta) \quad . \end{aligned} \tag{B-65}$$

The partial derivative with respect to β above reflects the fact that all implicit dependencies of $\mathscr{F}(\beta)$ on β through its dependence on other parameters, e.g., f_1, g_1, χ_1, μ_1, have vanishing derivatives since $\mathscr{F}(\beta)$ is evaluated at a saddle point. Making use of the saddle point equations, we find that

$$\beta \frac{\partial \left(\beta \mathscr{F}(\beta)\right)}{\partial \beta} = \chi_1 \left(1 - \frac{\beta}{\mu_1}\right) R(-\chi_1) + \beta \left(\frac{\chi_1}{\mu_1} + p_1 + q_1\right) R(-\chi_1 - \mu_1 p_1) - \beta q_1 R'(-\chi_1 - \mu_1 p_1). \tag{B-66}$$

Next, we need to analyze the behavior of $\beta \mathscr{F}(\beta)$ for large β. This can be seen directly through (B-59), (B-60). The first term in (B-59) can be shown to be of the form $\beta A + O(\beta^{-1})$, where $A \in \mathbb{R}$ is some constant. The reason for this behavior stems from the fact that for a discrete alphabet the corrections to the leading order term are exponential in β, except for a small $O(\beta^{-1})$ region close to the nearest neighbor points in the lattice. Therefore, to order β^{-1}, the first term in (B-59) is simply β times its partial derivative with respect to β. This enables us to evaluate the value of $\beta \mathscr{F}(\beta)$ for large β to the order β^{-1} as follows:

$$\begin{aligned} \beta \mathscr{F}(\beta) = \beta \frac{\partial}{\partial \beta} &\left[\frac{\beta}{\mu_1} \int \int_{\mathbb{C}} \log \left(\int_{\mathbb{C}} \left(\sum_{x \in \mathscr{B}_u} \mathrm{e}^{\, 2\beta \Re\{x(f_1 z^* + g_1 y^*)\} - \beta \varepsilon_1 |x|^2} \right)^{\frac{\mu_1}{\beta}} \mathrm{D}y \right) \mathrm{D}z \, \mathrm{d}F_U(u) \right] \\ &- \beta f_1^2 (\chi_1 + p_1 \mu_1) - \beta g_1^2 (\chi_1 + (p_1 + q_1)\mu_1) + \beta \varepsilon_1 (q_1 + p_1 + \frac{\chi_1}{\beta}) \\ - \left(1 - \frac{\beta}{\mu_1}\right) &\int_0^{\chi_1} R(-w) \, \mathrm{d}w - \frac{\beta}{\mu_1} \int_0^{\chi_1 + \mu_1 p_1} R(-w) \, \mathrm{d}w - \beta q_1 R(-\chi_1 - \mu_1 p_1) \quad . \end{aligned} \tag{B-67}$$

Using the fixed point equations we may re-express the first line as follows:

$$\begin{aligned} \frac{\beta}{\mu_1} \int_{\chi_1}^{\chi_1 + \mu_1 p_1} & R(-w) \mathrm{d}w + \beta q_1 R(-\chi_1 - \mu_1 p_1) + 2\beta q_1 R'(-\chi_1 - \mu_1 p_1) \\ & - \chi_1 R(-\chi_1) + \frac{2\beta \chi_1 R(-\chi_1)}{\mu_1} - 2\beta R(-\chi_1 - \mu_1 p_1) \left(\frac{\chi_1}{\mu_1} - q_1 - p_1\right) \quad . \end{aligned} \tag{B-68}$$

Plugging this into the above equation and using (B-45)–(B-47), we eventually get the following equation for the zero-temperature entropy

$$\bar{\mathscr{S}} \;=\; \chi_1 R(-\chi_1) - \int_0^{\chi_1} R(-w)\mathrm{d}w \quad . \tag{B-69}$$

Remarkably the above equation for the entropy holds also for the RS case. To recover the RS structure of the equations above we start with $\mu_1/\beta = 1$ and $\chi_0 = \chi_1 + \mu_1 p_1$. Then, we find that $q_0 = q_1$, $\epsilon_0 = \epsilon_1 - \beta g_1^2$, and $f_0 = f_1$. After that we find the equations to reduce to the RS case analyzed in [24].

C Proof of Proposition 3.1

We will now apply (3-9) to express the energy in a compact fashion. We start by considering the representation of the normalized average free energy in terms of $\boldsymbol{Q}$ (see (3-19) and (3-24)), and let us denote this representation, for the sake of clarity, as $\mathscr{F}(\boldsymbol{Q},\beta)$. In general, the replica crosscorrelation matrix $\boldsymbol{Q}$ depends on β. However, at the saddle point we have (by definition)

$$\frac{\partial \mathscr{F}(\boldsymbol{Q},\beta)}{\partial \boldsymbol{Q}} = \boldsymbol{0}. \tag{C-1}$$

Thus, the total derivative in (3-9) becomes a partial derivative at the saddle point, i.e.

$$\mathscr{E}(\beta) = \frac{\partial}{\partial \beta}\left(\beta \mathscr{F}(\boldsymbol{Q},\beta)\right) \quad . \tag{C-2}$$

Referring to the proof in Appendix B, then with (B-2), (B-5), (B-7), (B-11), and (B-15), while substituting $h = 0$, this gives

$$\mathscr{E}(\beta) = \lim_{n\to 0}\frac{1}{n}\frac{\partial}{\partial \beta}\int_0^{\beta} \mathrm{Tr}[\boldsymbol{Q}R(-w\boldsymbol{Q})]\,\mathrm{d}w \quad , \tag{C-3}$$

which is easily shown to be equivalent to (3-25). Furthermore, we get (3-26) by plugging (B-34) into (B-36) while substituting $h = 0$.

D Discrete Lattice Relaxation: Small χ_1 Approximation Near Unit Load (1RSB)

This appendix provides an approximate derivation of the 1RSB equations for the discrete lattice-based alphabet relaxation scheme of Section 6, while assuming a Gaussian $\boldsymbol{H}$, and a ZF front-end. The approximation is based on the numerical observation that the macroscopic parameter χ_1, employed in the 1RSB ansatz for this setting, approaches zero as the system load gets close to unity. This approximation considerably simplifies the numerical solution of the 1RSB equations in this region of the system load.

D.1 Case I: $\alpha = 1$

For $\alpha = 1$, the R-transform of $\boldsymbol{J} = \boldsymbol{T}^{\dagger}\boldsymbol{T}$ (see Proposition 4.1) satisfies

$$
\begin{aligned}
R(-w) &= \frac{\alpha - 1 + \sqrt{(1-\alpha)^2 + 4\alpha w}}{2\alpha w} \underset{\alpha=1}{=} \frac{1}{\sqrt{w}}, \quad \forall w \in \mathbb{R} && \text{(D-1)}\\
R'(-w) &= \frac{\left(1 - \alpha - \sqrt{(1-\alpha)^2 + 4\alpha w}\right)^2}{4\alpha w^2 \sqrt{(1-\alpha)^2 + 4\alpha w}} \underset{\alpha=1}{=} \frac{1}{2w^{\frac{3}{2}}}, \quad \forall w \in \mathbb{R} \quad . && \text{(D-2)}
\end{aligned}
$$

Considering the small χ_1 regime, we get from (4-2)–(4-4)

$$
\begin{aligned}
\varepsilon_1 &= \frac{1}{\sqrt{\chi_1}} \quad , && \text{(D-3)}\\
g_1 &= \sqrt{\frac{\frac{1}{\sqrt{\chi_1}} - \frac{1}{\sqrt{\chi_1 + \mu_1 p_1}}}{\mu_1}} \underset{\chi_1 \ll 1}{\approx} \sqrt{\frac{1}{\mu_1 \sqrt{\chi_1}}} \quad , && \text{(D-4)}\\
f_1 &\underset{\chi_1 \ll 1}{\approx} \sqrt{q_1 R'(-\mu_1 p_1)} = \sqrt{q_1 \frac{1}{2(\mu_1 p_1)^{\frac{3}{2}}}} \quad . && \text{(D-5)}
\end{aligned}
$$

Particularizing to the two-dimensional discrete lattice-based alphabet relaxation scheme in concern, one gets from (6-6)

$$
\psi_k(\xi) = \frac{\varepsilon_1 v_k - f_1 \xi}{g_1} \underset{\chi_1 \ll 1}{\approx} \frac{v_k}{\sqrt{\chi_1}} \sqrt{\mu_1 \sqrt{\chi_1}} = \sqrt{\mu_1} \frac{1}{\chi_1^{\frac{1}{4}}} \frac{c_k + c_{k-1}}{2} \quad \forall\, |\xi| < \infty. \tag{D-6}
$$

We now rewrite the function $\Theta_k(\xi)$ of (6-7) as

$$
\Theta_k(\xi) \triangleq \mathrm{e}^{\mu_1 c_k \left[(\mu_1 g_1^2 - \varepsilon) c_k + 2 f_1 \xi\right]} \left[Q\left(\sqrt{2}(\psi_k(\xi) - \mu_1 g_1 c_k)\right) - Q\left(\sqrt{2}(\psi_{k+1}(\xi) - \mu_1 g_1 c_k)\right)\right] \tag{D-7}
$$

and observe the following. Starting with exponential argument, we get

$$
(\mu_1 g_1^2 - \varepsilon_1) c_k + 2 f_1 \xi \underset{\chi_1 \ll 1}{\approx} -R(-\mu_1 p_1) c_k + 2 f_1 \xi = -\frac{c_k}{\sqrt{\mu_1 p_1}} + 2 f_1 \xi \quad , \tag{D-8}
$$

while the arguments of the $Q(\cdot)$ functions satisfiy

$$
\psi_k(\xi) - \mu_1 g_1 c_k \underset{\chi_1 \ll 1}{\approx} \frac{\sqrt{\mu_1}}{\chi_1^{\frac{1}{4}}} \frac{c_k + c_{k-1}}{2} - \frac{\sqrt{\mu_1}}{\chi_1^{\frac{1}{4}}} c_k = -\frac{\sqrt{\mu_1}}{\chi_1^{\frac{1}{4}}} \frac{c_k - c_{k-1}}{2} \quad , \tag{D-9}
$$

and

$$
\psi_{k+1}(\xi) - \mu_1 g_1 c_k \underset{\chi_1 \ll 1}{\approx} \frac{\sqrt{\mu_1}}{\chi_1^{\frac{1}{4}}} \frac{c_{k+1} + c_k}{2} - \frac{\sqrt{\mu_1}}{\chi_1^{\frac{1}{4}}} c_k = \frac{\sqrt{\mu_1}}{\chi_1^{\frac{1}{4}}} \frac{c_{k+1} - c_k}{2} \quad . \tag{D-10}
$$

Now recall that from the underlying definition of the extended alphabet set, it follows that $c_k \geq c_{k-1}\ \forall k$, and it can hence be concluded that

$$\psi_k(\xi) - \mu_1 g_1 c_k \underset{\chi_1 \ll 1}{\approx} \begin{cases} -\infty & c_{k-1} = -\infty, |c_k| < \infty \ , \\ -\frac{\sqrt{\mu_1}}{\chi_1^{\frac{1}{4}}} \frac{c_k - c_{k-1}}{2} \to -\infty & |c_{k-1}|\, , \ |c_k| < \infty \ , \end{cases} \tag{D-11}$$

and

$$\psi_{k+1}(\xi) - \mu_1 g_1 c_k \underset{\chi_1 \ll 1}{\approx} \begin{cases} \frac{\sqrt{\mu_1}}{\chi_1^{\frac{1}{4}}} \frac{c_{k+1} - c_k}{2} \to \infty & |c_k|\, , \ |c_{k+1}| < \infty \ , \\ \infty & c_k < \infty, |c_{k+1}| = \infty \ . \end{cases} \tag{D-12}$$

This implies that

$$Q(\sqrt{2}(\psi_k(\xi) - \mu_1 g_1 c_k)) \quad \xrightarrow{\chi_1 \to 0} \quad 1\ \forall k \tag{D-13}$$

$$Q(\sqrt{2}(\psi_{k+1}(\xi) - \mu_1 g_1 c_k)) \quad \xrightarrow{\chi_1 \to 0} \quad 0\ \forall k \ . \tag{D-14}$$

We therefore conclude that

$$\Theta_k(\xi) \underset{\chi_1 \ll 1}{\approx} \mathrm{e}^{\mu_1 c_k \left[(\mu_1 g_1^2 - \varepsilon_1) c_k + 2 f_1 \xi\right]} \underset{\chi_1 \ll 1}{\approx} \mathrm{e}^{\mu_1 c_k [2 f_1 \xi - R(-\mu_1 p_1) c_k]} = \mathrm{e}^{\mu_1 c_k \left(2 f \xi - \frac{c_k}{\sqrt{\mu_1 p_1}}\right)} \ . \tag{D-15}$$

In a similar manner one can observe that the exponential terms in the RHS of (6-8) vanish as $\chi_1 \to 0$, and conclude that

$$\Psi_k(\xi) \xrightarrow{\chi_1 \to 0} 0 \quad \forall k \ . \tag{D-16}$$

In view of the above we can now restate the coupled equations that determine the macroscopic parameters q_1, p_1, and μ_1 in the following way (cf. (6-9)–(6-12), and note that the equation for determining χ_1 can be ignored):

$$q_1 \underset{\chi_1 \ll 1}{\approx} 2 \int \frac{\sum_{m=1}^{L} c_m^2 \Theta_m(\xi)}{\sum_{m=1}^{L} \Theta_m(\xi)} \mathrm{e}^{-\xi^2} \frac{\mathrm{d}\xi}{\sqrt{\pi}} - p_1 \quad , \tag{D-17}$$

$$p_1 \underset{\chi_1 \ll 1}{\approx} \frac{2}{f_1 \mu_1} \int \frac{\sum_{m=1}^{L} c_m \Theta_m(\xi)}{\sum_{m=1}^{L} \Theta_m(\xi)} \xi \mathrm{e}^{-\xi^2} \frac{\mathrm{d}\xi}{\sqrt{\pi}} \ , \tag{D-18}$$

$$0 \underset{\chi_1 \ll 1}{\approx} 2 \int \log\left(\sum_{m=1}^{L} \Theta_m(\xi)\right) \mathrm{e}^{-\xi^2} \frac{\mathrm{d}\xi}{\sqrt{\pi}} \ , \tag{D-19}$$

where we used (D-1)–(D-5) to obtain

$$\int_0^{\mu_1 p_1} R(-w)\, \mathrm{d}w = 2\sqrt{\mu_1 p_1} \quad , \quad R'(-\mu_1 p_1) = \frac{1}{2(\mu_1 p_1)^{\frac{3}{2}}} \quad . \tag{D-20}$$

The energy penalty in this case is given by (cf. (4-12))

$$\bar{\mathscr{E}}_{\mathsf{rsb1}} \underset{\chi_1 \ll 1}{\approx} \frac{q_1 + p_1}{\sqrt{\mu_1 p_1}} - \frac{q_1 \mu_1 p_1}{2(\mu_1 p_1)^{\frac{3}{2}}} = \frac{q_1 + 2p_1}{2\sqrt{\mu_1 p_1}} \quad . \tag{D-21}$$

D.2 Case II: $\alpha < 1$, $\alpha \to 1$

In a similar manner to the previous section, we start with the R-transform of $\boldsymbol{J} = \boldsymbol{T}^{\dagger}\boldsymbol{T}$, and rewrite it for small w, using the Taylor expansion around $w = 0$, as

$$\begin{aligned} R(-w) &= \frac{\alpha - 1 + \sqrt{(1-\alpha)^2 + 4\alpha w}}{2\alpha w} \underset{w \ll 1}{\approx}, \frac{-(1-\alpha) + (1-\alpha)\left(1 + \frac{2\alpha}{(1-\alpha)^2}w - \frac{2\alpha^2}{(1-\alpha)^4}w^2 + o(w^2)\right)}{2\alpha w} \\ &\underset{w \ll 1}{\approx} \frac{1}{1-\alpha} - \frac{\alpha}{(1-\alpha)^3}w + O(w^2) \ . \end{aligned} \tag{D-22}$$

We focus in the following on the regime in which $\alpha \to 1$, so that $\frac{1}{1-\alpha} \gg 1$, but still $\frac{1}{1-\alpha} \ll \frac{1}{\chi_1}$. It hence follows that

$$\varepsilon_1 \underset{\chi_1 \ll 1}{\approx} \frac{1}{1-\alpha} + O(\chi_1) \quad , \tag{D-23}$$

$$g_1 \underset{\chi_1 \ll 1}{\approx} \sqrt{\frac{\frac{1}{1-\alpha} - \frac{1}{\sqrt{\mu_1 p_1}}}{\mu_1}} \underset{\chi_1 \ll 1, \alpha \to 1}{\approx} \sqrt{\frac{1}{\mu_1(1-\alpha)}} \quad , \tag{D-24}$$

$$f_1 \underset{\chi_1 \ll 1}{\approx} \sqrt{q_1 R'(-\mu_1 p_1)} \quad . \tag{D-25}$$

Particularizing again to the two-dimensional discrete lattice alphabet relaxation scheme for QPSK signaling, it follows from (6-6) that

$$\psi_k(\xi) = \frac{\varepsilon_1 v_k - f_1 \xi}{g_1} \underset{\chi_1 \ll 1, \alpha \to 1}{\approx} \frac{v_k}{1-\alpha}\sqrt{\mu_1(1-\alpha)} = \sqrt{\frac{\mu_1}{1-\alpha}}\frac{c_k + c_{k-1}}{2} \quad \forall\, |\xi| < \infty \quad . \tag{D-26}$$

Considering (6-7) we write

$$(\mu_1 g_1^2 - \varepsilon_1)c_k + 2f_1\xi \underset{\chi_1 \ll 1}{\approx} -R(-\mu_1 p_1)c_k + 2f_1\xi \quad . \tag{D-27}$$

Next, the arguments of the $Q(\cdot)$ functions in (6-7) satisfy

$$\psi_k(\xi) - \mu_1 g_1 c_k \underset{\chi_1 \ll 1, \alpha \to 1}{\approx} \sqrt{\frac{\mu_1}{1-\alpha}}\frac{c_k + c_{k-1}}{2} - \sqrt{\frac{\mu_1}{1-\alpha}}c_k = -\sqrt{\frac{\mu_1}{1-\alpha}}\frac{c_k - c_{k-1}}{2} \quad , \tag{D-28}$$

and

$$\psi_{k+1}(\xi) - \mu_1 g_1 c_k \underset{\chi_1 \ll 1, \alpha \to 1}{\approx} \sqrt{\frac{\mu_1}{1-\alpha}}\frac{c_{k+1} + c_k}{2} - \sqrt{\frac{\mu_1}{1-\alpha}}c_k = \sqrt{\frac{\mu_1}{1-\alpha}}\frac{c_{k+1} - c_k}{2} \quad . \tag{D-29}$$

This enables us to conclude that

$$\psi_k(\xi) - \mu_1 g_1 c_k \xrightarrow{\chi_1 \ll 1, \alpha \to 1} -\infty \tag{D-30}$$

$$\psi_{k+1}(\xi) - \mu_1 g_1 c_k \xrightarrow{\chi_1 \ll 1, \alpha \to 1} \infty \ , \tag{D-31}$$

and hence

$$\Theta_k(\xi) \underset{\chi_1 \ll 1, \alpha \to 1}{\approx} \mathrm{e}^{\,\mu_1 c_k \left[(\mu_1 g_1^2 - \varepsilon_1) c_k + 2 f_1 \xi\right]} \underset{\chi_1 \ll 1, \alpha \to 1}{\approx} \mathrm{e}^{\,\mu_1 c_k [2 f_1 \xi - R(-\mu_1 p_1) c_k]} \;, \quad \forall k \quad , \tag{D-32}$$

and

$$\Psi_k(\xi) \xrightarrow{\chi_1 \ll 0, \alpha \to 1} 0 \quad , \quad \forall k \; . \tag{D-33}$$

Finally, note that $\int_0^{\mu_1 p_1} R(-w)\, dw$ exists for $\alpha < 1$, and the approximation

$$\mu_1 \chi_1 g_1^2 \xrightarrow{\chi_1 \ll 0, \alpha \to 1} 0 \tag{D-34}$$

is employed to derive the three coupled equation that determine the macroscopic parameters q_1, p_1, and μ_1. The three equations are thus

$$q_1 \quad \underset{\chi_1 \ll 1, \alpha \to 1}{\approx} \quad 2 \int \frac{\sum_{m=1}^{L} c_m^2 \Theta_m(\xi)}{\sum_{m=1}^{L} \Theta_m(\xi)} \mathrm{e}^{\,-\xi^2} \frac{\mathrm{d}\xi}{\sqrt{\pi}} - p_1 \quad , \tag{D-35}$$

$$p_1 \quad \underset{\chi_1 \ll 1, \alpha \to 1}{\approx} \quad \frac{2}{f_1 \mu_1} \int \frac{\sum_{m=1}^{L} c_m \Theta_m(\xi)}{\sum_{m=1}^{L} \Theta_m(\xi)} \xi \mathrm{e}^{\,-\xi^2} \frac{\mathrm{d}\xi}{\sqrt{\pi}} \; , \tag{D-36}$$

$$\mu_1 \quad \underset{\chi_1 \ll 1, \alpha \to 1}{\approx} \quad \frac{2 \int \log \left(\sum_{m=1}^{L} \Theta_m(\xi) \right) \mathrm{e}^{\,-\xi^2} \frac{\mathrm{d}\xi}{\sqrt{\pi}} - \int_0^{\mu_1 p_1} R(-w)\, \mathrm{d}w + \mu_1 (q_1 + 2 p_1) R(-\mu_1 p_1)}{2 q_1 \mu_1 p_1 R'(-\mu_1 p_1)} \; . \tag{D-37}$$

The expression for the energy penalty is given by

$$\bar{\mathscr{E}}_{\mathsf{rsb1}} \underset{\chi_1 \ll 1, \alpha \to 1}{\approx} (q_1 + p_1)\, R(-\mu_1 p_1) - q_1 \mu_1 p_1 R'(-\mu_1 p_1) \; . \tag{D-38}$$

We also note that the exact expressions for the R-transform and its derivative were employed for the purpose of producing more accurate numerical results, while using this small χ_1 approximation for $\alpha < 1$.

E Proof of Lemma 4.6

The Stieltjes transform of the probability distribution $F(x)$ is defined by

$$m(s) = \int \frac{\mathrm{d}F(x)}{x - s} \quad . \tag{E-1}$$

In terms of the Stieltjes transform, the R-transform is defined as

$$R(w) = m^{-1}(-w) - \frac{1}{w} \quad , \tag{E-2}$$

where $m^{-1}(s)$ denotes the inverse function of $m(s)$ with respect to composition, i.e., $m(m^{-1}(s)) = s$.

We start with the observation that the derivative of the Stieltjes transform is lower bounded by its square

$$m'(s) = \int \frac{\mathrm{d}F(x)}{(x-s)^2} \geq [m(s)]^2 \quad , \tag{E-3}$$

by means of Jensen's inequality, with equality if and only if the distribution $F(x)$ is a single mass point. Next, we consider the derivative of the R-transform. Letting $w = m(s)$, it follows that

$$\begin{aligned} R'(w) &= \frac{\mathrm{d}m^{-1}(-w)}{\mathrm{d}w} + \frac{1}{w^2} & \text{(E-4)} \\ &= \frac{-1}{m'(s)} + \frac{1}{[m(s)]^2} \geq 0 \quad , & \text{(E-5)} \end{aligned}$$

with equality if and only if the distribution $F(x)$ is a single mass point. Lemma 4.6 then follows immediately.

F Spectral Efficiency of Generalized Tomlinson-Harashima Precoding

For the sake of comparison, we review here the derivation of the spectral efficiency of *generalized Tomlinson-Harashima precoding (GTHP)*, which is another practical alternative to the capacity achieving DPC. The approach is based on inflated lattice strategies, and borrows ideas from the recent analysis of pulse amplitude modulation (PAM) in [54]. The spectral efficiency is derived following [6, 55] (see also [20, 56]), while employing successive encoding using the inflated lattice strategy at each stage, where the signals of previously encoded users are treated as *causally* known interference. We consider here the "canonical" channel model as in (2-1), and note that a comparative analysis of other variants of GTHP can be found, e.g., in [20].

The underlying idea of the scheme considered here is first to induce a "triangular" channel structure using the LQ-factorization of the channel transfer matrix. Assuming $\boldsymbol{H}$ is full rank, we denote

$$\boldsymbol{H} = \boldsymbol{L}\bar{\mathbf{Q}} \quad , \tag{F-1}$$

where $\boldsymbol{L}_{[K\times K]}$ is lower triangular with positive diagonal entries and $\bar{\mathbf{Q}}_{K\times N}$ has orthonormal rows. The transmitted signal is then given by

$$\boldsymbol{t} = \bar{\mathbf{Q}}^{\dagger}\boldsymbol{x} \quad , \tag{F-2}$$

where $\boldsymbol{x}$ is the nonlinear precoder's output (cf. (2-3)). The signal received by the kth user is thus given by

$$r_k = L_{kk}x_k + \sum_{j=1}^{k-1} L_{kj}x_j + n_k \quad , \tag{F-3}$$

where $\{L_{ij}\}$ denote the entries of $\boldsymbol{L}$ and x_i is the nonlinear precoder's output that corresponds to user i. Normalizing both sides of the equation by L_{kk}, we get the following equivalent channel

$$\begin{aligned}\breve{r}_k &= x_k + \sum_{j=1}^{k-1} \frac{L_{kj}}{L_{kk}} x_j + \breve{n}_k \\ &= x_k + s_k + \breve{n}_k \quad ,\end{aligned} \tag{F-4}$$

where we denote the multiuser interference experienced by user k as $s_k \triangleq \sum_{j=1}^{k-1} \frac{L_{kj}}{L_{kk}} x_j$, and $\breve{n}_k$ is a zero-mean circularly symmetric complex Gaussian noise with variance $\frac{\sigma^2}{L_{kk}^2}$.

In the GTHP setting, instead of DPC (as employed at this point, e.g., by the "zero-forcing dirty-paper" scheme of [49]), we follow for each user the THP-type strategy described in [55] for canceling the interference due to previously encoded users. This strategy, which applies for canceling *causally* known interference, leads in the broadcast setting to a considerably reduced complexity as it involves only scalar quantizations (as opposed to vector quantizations in the noncausal case, see therein). To make a fair comparison to the precoding schemes discussed in Sections 6-7, we particularize here to the case in which the information bearing signal takes on binary values per each dimension (so that the total spectral efficiency for quadrature modulation, as a function of $\frac{E_b}{N_0}$, is twice as much as the one obtained for binary input). The spectral efficiency for continuous input is derived as well for completeness. The basic transmission scheme is reviewed first, while considering real channels.

The underlying *real* channel model is given by

$$y = x + s + n \quad , \tag{F-5}$$

where x is subject to an average power constraint P_x, n is a zero-mean AWGN with variance P_n, and s is an interference signal which is known causally at the transmitter (i.e., at the current time instance), but not at the receiver. This channel model is also referred to in the literature as the "dirty-tape" model [55]. Consider the one-dimensional lattice

$$\Lambda = \Delta \{\cdots - 3, -1, 1, 3, \cdots\} \ . \tag{F-6}$$

Let $\mathcal{V} = [-\Delta, \Delta)$ denote the basic Voronoi region of Λ. Let d be a dither signal uniformly distributed over $\mathcal{V}$. Under a common randomness assumption, this dither signal is assumed to be available at the receiver as well.

Starting with *continuous* information bearing signals, then by the GTHP scheme the transmitter sends the signal

$$x = [v - \tilde{\alpha} s - d] \mod \Lambda \quad , \tag{F-7}$$

where $\tilde{\alpha} \in (0, 1]$ is referred to as the "inflation factor". The receiver scales the received signal by $\tilde{\alpha}$, adds the dither signal, and then performs a modulo-lattice operation, yielding

$$y' = [\tilde{\alpha} y + d] \mod \Lambda \quad . \tag{F-8}$$

Effectively, the induced channel is equivalent to (see [55], Lemma 6)

$$y' = [v + n_{\text{eff}}] \mod \Lambda \quad , \tag{F-9}$$

where the effective noise is given by

$$n_{\text{eff}} \triangleq [(\tilde{\alpha} - 1)x + \tilde{\alpha}n] \mod \Lambda \quad , \tag{F-10}$$

and we note that the dither signal ensures that x is uniformly distributed over the Voronoi region, and is independent of either the information bearing signal v, or the noise n.

The capacity of this channel is achieved by a uniform input distribution over the Voronoi region, $v \sim \mathsf{Unif}\{\mathcal{V}\}$, for which the relation between the lattice constant Δ and the transmit power P_x is given by $\Delta = \sqrt{3P_x}$. The corresponding achievable rate is equal to the input-output mutual information of the equivalent channel (F-9)

$$\begin{aligned} R(P_x) &\triangleq \mathrm{I}(v, y') = \log(2\Delta) - \mathrm{h}(n_{\text{eff}}) \\ &= \frac{1}{2}\log(12P_x) - \int_{-\Delta}^{\Delta} f_{n_{\text{eff}}}(\zeta)\log_2 f_{n_{\text{eff}}}(\zeta)\,\mathrm{d}\zeta \Bigg|_{\Delta=\sqrt{3P_x}} \quad . \end{aligned} \tag{F-11}$$

The entropy of the effective noise is derived via the following observation. Denoting the "self-noise" term by

$$Z = (\tilde{\alpha} - 1)x \quad , \tag{F-12}$$

its pdf is given by

$$f_Z(\zeta) = \begin{cases} \frac{1}{2(1-\tilde{\alpha})\Delta} & |\zeta| \le (1-\tilde{\alpha})\Delta \ , \\ 0 & \text{otherwise} \ . \end{cases} \tag{F-13}$$

The pdf of the effective noise (F-10) is thus given by

$$f_{n_{\text{eff}}}(\zeta) = \begin{cases} \sum_{i=-\infty}^{\infty} f_{\tilde{Z}}(\zeta - 2i\Delta) & -\Delta \le \zeta < \Delta \ , \\ 0 & \text{otherwise} \ , \end{cases} \tag{F-14}$$

where $f_{\tilde{Z}}(\zeta)$ denotes the pdf of the pre-modulo noise term, which is given by the convolution of the pdf of the self-noise (F-13) and the pdf of the scaled AWGN

$$\begin{aligned} f_{\tilde{Z}}(\zeta) &= f_Z(\zeta) * f_{\tilde{\alpha}n}(\zeta) \\ &= \frac{1}{2(1-\tilde{\alpha})\Delta}\left[Q\left(\frac{\sqrt{2}}{\tilde{\alpha}}(\zeta - (1-\tilde{\alpha})\Delta)\right) - Q\left(\frac{\sqrt{2}}{\tilde{\alpha}}(\zeta + (1-\tilde{\alpha})\Delta)\right)\right] \quad . \end{aligned} \tag{F-15}$$

We normalized here without loss of generality the spectral level of the AWGN to $\frac{1}{2}$ per dimension (inducing a unit noise spectral level in complex channels [50]), so that effectively P_x specifies the SNR of the original underlying *complex* channel model, corresponding to (F-4). The rate in (F-11) can be optimized with respect to the inflation factor $\tilde{\alpha}$ (which is performed to obtain the numerical results shown in Section 8). We also note that choosing $\tilde{\alpha} = 1$ corresponds to standard THP, while another popular choice is the minimum mean-squared error (MMSE) factor (also referred to as the "Costa factor" [4])

$$\alpha_{\mathrm{MMSE}} = \frac{P_x}{P_x + P_n} \quad . \tag{F-16}$$

Turning to discrete input with *M-pulse amplitude modulation (M-PAM)* (representing the information bearing signals), the setting is equivalent to the case in which the continuous information bearing signal considered above is *quantized* (cf. [57]). Instead of (F-7), the channel input is now given by

$$x = [\mathcal{Q}(v) - \tilde{\alpha}s - d] \mod \Lambda \triangleq [v_{\mathcal{Q}} - \tilde{\alpha}s - d] \mod \Lambda \quad , \tag{F-17}$$

where $\mathcal{Q}(\cdot)$ denotes the nearest-neighbor uniform quantizer with step size Δ [54], and v is assumed to be uniformly distributed over the Voronoi region. We note here that this transmission scheme differs from the one considered in [54], where the *channel input* is quantized to comply with an M-PAM constellation (see therein). Note also that as in the continuous setting, due to the dither signal, the channel input x is still uniformly distributed over the Voronoi region. The effective channel can now be represented in the form (cf. (F-9))

$$y'_{\mathcal{Q}} = [v_{\mathcal{Q}} + n_{\mathrm{eff}}] \mod \Lambda \quad , \tag{F-18}$$

where the effective noise is still given by (F-10). Restricting this review to the case of binary information bearing signals per dimension, the channel input signal is limited to the interval $\mathcal{V} = [-\Delta, \Delta)$, while the quantized information bearing signal is obtained using

$$\mathcal{Q}(v) = \begin{cases} -\frac{\Delta}{2} & -\Delta \leq v < 0 \ , \\ +\frac{\Delta}{2} & 0 \leq v < \Delta \ . \end{cases} \tag{F-19}$$

For consistency we retain the relation $\Delta = \sqrt{3P_{x_{\mathcal{Q}}}}$.

The achievable rate for binary input is given again by the mutual information

$$R(P_{x_{\mathcal{Q}}}) \triangleq \mathrm{I}(v; y'_{\mathcal{Q}}) = \mathrm{h}(y'_{\mathcal{Q}}) - \mathrm{h}(n_{\mathrm{eff}}) \quad . \tag{F-20}$$

Note that the pdf of the random quantity inside the modulo function in (F-18) is given by

$$\begin{aligned} f_{\tilde{Y}}(\zeta) &= f_{v_{\mathcal{Q}}}(\zeta) * f_{\tilde{Z}}(\zeta) \\ &= \left(\frac{1}{2}\delta\left(\zeta - \frac{\Delta}{2}\right) + \frac{1}{2}\delta\left(\zeta + \frac{\Delta}{2}\right)\right) * f_{\tilde{Z}}(\zeta) \\ &= \frac{1}{2}f_{\tilde{Z}}\left(\zeta - \frac{\Delta}{2}\right) + \frac{1}{2}f_{\tilde{Z}}\left(\zeta + \frac{\Delta}{2}\right) \quad . \end{aligned} \tag{F-21}$$

Hence, the pdf of the equivalent channel output $y'_{\mathcal{Q}}$ is equal to

$$f_{y'_{\mathcal{Q}}}(\zeta) = \begin{cases} \sum_{i=-\infty}^{\infty} f_{\tilde{Y}}(\zeta - 2i\Delta) & \Delta \le \zeta < \Delta \ , \\ 0 & \text{otherwise} \ , \end{cases} \tag{F-22}$$

and the achievable rate of (F-20) can be rewritten as

$$\begin{aligned} R(P_{x_{\mathcal{Q}}}) &= -\int_{-\Delta}^{\Delta} f_{y'_{\mathcal{Q}}}(\zeta) \log_2 f_{y'_{\mathcal{Q}}}(\zeta)\, \mathrm{d}\zeta + \int_{-\Delta}^{\Delta} f_{n_{\mathrm{eff}}}(\zeta) \log_2 f_{n_{\mathrm{eff}}}(\zeta)\, \mathrm{d}\zeta \\ &= \int_{-\Delta}^{\Delta} \Big[f_{n_{\mathrm{eff}}}(\zeta) \log_2 f_{n_{\mathrm{eff}}}(\zeta) - f_{y'_{\mathcal{Q}}}(\zeta) \log_2 f_{y'_{\mathcal{Q}}}(\zeta) \Big]\, \mathrm{d}\zeta \Bigg|_{\Delta = \sqrt{3P_{x_{\mathcal{Q}}}}} \quad . \end{aligned} \tag{F-23}$$

The above principles can now be applied to the channel in (F-3), where the transmitter precancells using the GTHP scheme, per each transmitted symbol, the interference due to the *corresponding symbols* of previously encoded users. Using (F-11) and (F-20), the achievable rate of the kth user can be obtained by substituting $P_x = L_{kk}^2\, \mathsf{snr}$ for continuous input, and $P_{x_{\mathcal{Q}}} = L_{kk}^2\, \mathsf{snr}$ for the binary setting, yielding, respectively, for *real* channels

$$R_{C,k}^{\mathsf{gthp}}(\mathsf{snr}) = \frac{1}{2} \log(12 L_{kk}^2 \mathsf{snr}) + \int_{-\Delta}^{\Delta} f_{n_{\mathrm{eff}}^{C}}(\zeta) \log_2 f_{n_{\mathrm{eff}}^{C}}(\zeta)\, \mathrm{d}\zeta \Bigg|_{\Delta = \sqrt{3 L_{kk}^2 \mathsf{snr}}} \quad , \tag{F-24}$$

and

$$R_{\mathcal{Q},k}^{\mathsf{gthp}}(\mathsf{snr}) = \int_{-\Delta}^{\Delta} \Big[f_{n_{\mathrm{eff}}}(\zeta) \log_2 f_{n_{\mathrm{eff}}}(\zeta) - f_{y'_{\mathcal{Q}}}(\zeta) \log_2 f_{y'_{\mathcal{Q}}}(\zeta) \Big]\, \mathrm{d}\zeta \Bigg|_{\Delta = \sqrt{3 L_{kk}^2\, \mathsf{snr}}} \quad . \tag{F-25}$$

To complete the analysis, it is left to derive the normalized spectral efficiency of GTHP in the large system limit. This is obtained using the following observation (see [49, Lemma 3]).

Lemma F.1 *Let $\boldsymbol{H}$ be a $K \times N$ random matrix, having i.i.d. circularly symmetric zero-mean entries with variance $\frac{1}{N}$ and finite fourth moment, and let $\boldsymbol{H}^{(k)}$, $k < K$, denote the matrix constructed by striking out the* last *$K - k$ rows of $\boldsymbol{H}$. Then*

$$L_{kk}^2 = \frac{1}{\Big[(\boldsymbol{H}^{(k)} \boldsymbol{H}^{(k)\dagger})^{-1} \Big]_{kk}} \quad , \tag{F-26}$$

and for $k, K, N \to \infty$, s.t. $\frac{K}{N} \to \alpha < \infty$ and $\frac{k}{K} \to \nu \in [0, 1)$, it follows that

$$L_{kk}^2 \xrightarrow[k,K,N \to \infty]{} L^2(\nu) \triangleq 1 - \nu\alpha \ , \quad \alpha \in (0, 1] \quad . \tag{F-27}$$

Omitting subscripts, the limiting spectral efficiency is thus given for either continuous or binary quantized input by (cf. [49, Eq. (41)])

$$
\begin{aligned}
C^{\mathsf{gthp}}(\mathsf{snr}) &= \lim_{K,N\to\infty} \frac{1}{N}\sum_{k=1}^{K} R_k^{\mathsf{gthp}}(\mathsf{snr}) && \text{(F-28)}\\
&= \lim_{K,N\to\infty} \frac{K}{N}\frac{1}{K}\sum_{k=1}^{K} R_k^{\mathsf{gthp}}(\mathsf{snr}) && \text{(F-29)}\\
&= \alpha\int_0^1 R^{\mathsf{gthp}}(\nu,\mathsf{snr})\,\mathrm{d}\nu \quad , && \text{(F-30)}
\end{aligned}
$$

where for the continuous case we substitute

$$
R^{\mathsf{gthp}}(\nu,\mathsf{snr}) = R_C^{\mathsf{gthp}}(\nu,\mathsf{snr}) \triangleq \frac{1}{2}\log_2(12L^2(\nu)\mathsf{snr}) + \int_{-\Delta}^{\Delta} f_{n_{\mathrm{eff}}}(\zeta)\log_2 f_{n_{\mathrm{eff}}}(\zeta)\,\mathrm{d}\zeta\Bigg|_{\Delta=\sqrt{3L^2(\nu)\mathsf{snr}}} \quad , \tag{F-31}
$$

and for the case of discrete binary information bearing signals we substitue

$$
R^{\mathsf{gthp}}(\nu,\mathsf{snr}) = R_Q^{\mathsf{gthp}}(\nu,\mathsf{snr}) \triangleq \int_{-\Delta}^{\Delta}\Big[f_{n_{\mathrm{eff}}}(\zeta)\log_2 f_{n_{\mathrm{eff}}}(\zeta) - f_{y'_Q}(\zeta)\log_2 f_{y'_Q}(\zeta)\Big]\,\mathrm{d}\zeta\Bigg|_{\Delta=\sqrt{3L^2(\nu)\,\mathsf{snr}}} \quad . \tag{F-32}
$$

The spectral efficiency for QPSK modulation satisfies (following the convention in [50]):

$$
C_{\mathsf{qpsk}}^{\mathsf{gthp}}(\mathsf{snr}) = 2C_{\mathsf{bpsk}}^{\mathsf{gthp}}(\mathsf{snr}/2) \quad , \tag{F-33}
$$

where $C_{\mathsf{bpsk}}^{\mathsf{gthp}}(\mathsf{snr})$ is given by (F-30) and (F-32), and it can be expressed as a function of $\frac{E_b}{N_0}$ through (5-9). An analogous result for the case of continuous input can be readily obtained using (F-31). Both spectral efficiencies can be optimized with respect to the choice of the system load α.

References

[1] G. Caire, S. Shamai (Shitz), Y. Steinberg, and H. Weingarten, “On information-theoretic aspects of MIMO broadcast channels,” in *Space-Time Wireless Systems: From Array Processing to MIMO Communications*, H. Bölcskei, D. Gesbert, C. B. Papadias, and A.-J. van der Veen, Eds. Cambridge University Press, 2006, ch. 19.

[2] D. Astély, E. Dahlman, A. Furuskär, Y. Jading, M. Loindström, and S. Parkvall, “LTE: The evolution of mobile broadband,” *IEEE Communications Magazine*, vol. 47, no. 4, pp. 44–51, April 2009.

[3] Q. Lin, X. E. Lin, J. Zhang, and W. Roh, “Advancement of MIMO technology in WiMAX: From IEEE 802.16d/e/j to 802.16m,” *IEEE Communications Magazine*, vol. 47, no. 6, pp. 100–107, June 2009.

[4] M. H. M. Costa, “Writing on dirty paper,” *IEEE Transactions on Information Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[5] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, September 2006.

[6] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattic codes for structured multi-terminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

[7] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3417–3432, October 2005.

[8] A. Bennatan, D. Burstein, G. Caire, and S. Shamai (Shitz), "Superposition coding for dirty paper channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1872–1889, May 2006.

[9] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communications – Part II: Perturbation," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 537–544, March 2005.

[10] R. F. Fischer, *Precoding and Signal Shaping for Digital Transmission.* John Wiley & Sons, 2002.

[11] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electron. Lett.*, vol. 7, pp. 138–139, March 1971.

[12] H. Harashima and H. Miyakawa, "Matched-transmission technique fo channels with intersymbol interference," *IEEE Transactions on Communications*, vol. COM-20, no. 4, pp. 774–780, August 1972.

[13] D. A. Schmidt, M. Joham, and W. Utschick, "Minimum mean square error vector precoding," *European Transactions on Telecommunications*, vol. 19, pp. 219–131, 2008.

[14] M. Payaró and D. P. Palomar, "On optimal precoding in linear vector Gaussian channels with arbitrary input distribution," in *Proceedings of the 2009 IEEE International Symposium on Information Theory (ISIT'09)*, Seoul, Korea, June 28 – July 3, 2009.

[15] E. Arell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2201–2214, August 2002.

[16] C. Windpassinger, R. F. H. Fischer, and J. B. Huber, "Lattice-redcution-aided broadcast precoding," *IEEE Transactions on Communications*, vol. 52, no. 12, pp. 2057–2060, December 2004.

[17] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "Communication over MIMO broadcast channels using lattice-basis reduction," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4567–4582, December 2007.

[18] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, “Precoding in multiantenna and multiuser communications,” *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1305–1316, July 2004.

[19] M. Stojnic, H. Vikalo, and B. Hassibi, “Asymptotic analysis of the Gaussian broadcast channel with perturbation preprocessing,” in *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France, May 14–19, 2006.

[20] F. Boccardi, F. Tosato, and G. Caire, “Precoding schemes for the MIMO-GBC,” in *Proceedings of International Zurich Seminar on Communication*, Zurich, Switzerland, Feb. 2006.

[21] D. J. Ryan, I. B. Collings, I. V. L. Clarkson, and R. W. Heath Jr., “Performance of vector perturbation multiuser MIMO systems with limited feedback,” *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2633–2644, September 2009.

[22] A. Mobasher and A. K. Khandani, “Precoding in multiple-antenna broadcast systems with a probabilistic viewpoint,” in *Proceedings of the Canadian Workshop on Information Theory (CWIT’07)*, Edmonton, AB, Canada, June 6–8, 2007, pp. 132–135.

[23] A. Mobasher, M. A. Maddah-Ali, and A. K. Khandani, “Selective mapping for channel inversion precoding in multiple-antenna broadcast systems,” in *Proceedings of the 2009 IEEE International Symposium on Information Theory (ISIT’09)*, Seoul, Korea, June 28 – July 3, 2009.

[24] R. R. Müller, D. Guo, and A. L. Moustakas, “Vector precoding in high dimensions: A replica analysis,” *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 3, pp. 530–540, April 2008.

[25] R. de Miguel and R. R. Müller, “On convex vector precoding for multiuser MIMO broadcast channels,” *IEEE Trans. Sign. Proc.*, vol. 57, no. 11, pp. 4497–4508, November 2009.

[26] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction.* Oxford University Press, 2001.

[27] M. Mézard and A. Montanari, *Information, Physics, and Computation.* Oxford University Press, 2009.

[28] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors,” *IEEE Transactions on Information Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[29] R. R. Müller and W. H. Gerstacker, “On the capacity loss due to separation of detection and decoding,” *IEEE Transactions on Information Theory*, vol. 50, no. 8, pp. 1769–1778, Aug. 2004.

[30] D. Guo and S. Verdú, “Randomly spread CDMA: Asymptotics via statistical physics,” *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.

[31] D. Guo and T. Tanaka, “Generic multiuser detection and statistical physics,” in *Advances in Multiuser Detection*, M. L. Honig, Ed. Wiley-IEEE Press, 2009.

[32] D. Sherrington and S. Kirkpatrick, “Solvable model of a spin glass,” *Phys. Rev. Lett.*, vol. 35, pp. 1792–1796, 1972.

[33] L. A. Pastur and M. V. Shcherbina, “Absence of self-averaging of the order parameter in the Sherrington-Kirkpatrick model,” *Journal of Statistical Physics*, vol. 62, no. 1/2, pp. 1–19, 1991.

[34] F. Guerra and F. L. Toninelli, “The thermodynamic limit in mean field spin glass models,” *Commun. Math. Phys.*, vol. 230, pp. 71–79, 2002.

[35] F. Guerra, “The infinite volume limit in generalized mean field disordered models,” *Markov Proc. Rel. Fields*, vol. 9, pp. 195–207, 2003.

[36] S. Korada and N. Macris, “Tight bounds on the capacity of binary input random CDMA systems,” *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5590–5613, November 2010.

[37] M. Talagrand, “The Parisi formula,” *Annals of Mathematics*, vol. 163, no. 1, pp. 221–263, 2006.

[38] G. Parisi, “A sequence of approximate solutions to the S-K model for spin glasses,” *J. Phys. A: Math. Gen.*, vol. 13, 1980.

[39] F. Guerra, “Replica broken bounds in the mean field spin glass model,” *Commun. Math. Phys.*, vol. 233, pp. 1–12, 2003.

[40] M. L. Mehta, *Random Matrices*, 2nd ed. Boston, MA: Academic Press, 1991.

[41] D. N. C. Tse, “Multiuser receivers, random matrices and free probability,” in *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, Sep. 22 – 24, 1999, pp. 1055–1064.

[42] R. de Miguel, V. Gardašević, R. R. Müller, and F. Knudsen, “On overloaded vector precoding for single-user MIMO channels,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 745–753, Feb. 2010.

[43] S. Shamai (Shitz) and S. Verdú, “The impact of frequency-flat fading on the spectral efficiency of CDMA,” *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.

[44] P. Viswanath, D. N. C. Tse, and V. Anantharam, “Asymptotically optimal waterfilling in vector multiple access channels,” *IEEE Transactions on Information Theory*, vol. 47, no. 1, pp. 241–267, Jan. 2001.

[45] B. M. Zaidel, S. Shamai (Shitz), and S. Verdú, "Multi-cell uplink spectral efficiency of coded DS-CDMA with random signatures," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 8, pp. 1556–1569, Aug. 2001.

[46] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.

[47] S. Verdú, *Multiuser Detection.* Cambridge, UK: Cambridge University Press, 1998.

[48] R. R. Müller, "Multiuser receivers for randomly spread signals: Fundamental limits with and without decision-feedback," *IEEE Transactions on Information Theory*, vol. 47, no. 1, pp. 268–283, Jan. 2001.

[49] G. Caire and S. Shamai (Shitz), "On the achievable throughput of multiantenna Gaussian broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[50] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1329–1343, Jun. 2002.

[51] M. Mezard, "Personal communication," 2009.

[52] A. L. Moustakas and S. H. Simon, "On the outage capacity of correlated multiple-path MIMO channels," *IEEE Trans. Inform. Theory*, vol. 53, no. 11, pp. 3887–3903, November 2007.

[53] A. Guionnet and M. Maïda, "A Fourier view on the R-tansform and related asymptotics of spherical integrals," *Journal of Functional Analysis*, vol. 222, no. 2, pp. 435–490, 2005.

[54] T. Gariby, U. Erez, and S. Shamai (Shitz), "Dirty paper coding for PAM signaling," in *Proceedings of the 2007 IEEE International Symposium on Information Theory (ISIT'2007)*, Nice, France, June 24 – 29, 2007.

[55] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005, see also: — "Capacity and Lattice-Strategies for Cancelling Known Interference," Proceedings of ISITA'2000, Honolulu, HI, Aug. 2000.

[56] G. Ginis and J. M. Cioffi, "Vectored transmission for digital subscriber line systems," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1085–1104, June 2002.

[57] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, October 2004.